nature

CLEAR CONNECTIONS

Architecture of physical networks rendered in 3D **PAGE 676**

RESEARCH

BULLYING IN SCIENCE

Are universities and funding agencies doing enough?

PAGE 616

MICROBIOLOGY

A ROUTE TO RESISTANCE

Virus helps Staphylococcus evade the immune system

PAGES 637 & 705

PALAEOANTHROPOLOGY

NEANDERTHAL VIOLENCE

Skulls reveal hominins' bad reputation is unwarranted

PAGES 634 & 686

THIS WEEK

Beware the rise of the radical right

Academic freedom is on the hit list when politicians of the extreme right gain office — as they have done in some European countries.

Hidden inside a 1970s office block close to London's Waterloo station is a tiny organization that has helped tens of thousands of academics find sanctuary from conflict. Co-founded 85 years ago by the economist William Beveridge and physicist Ernest Rutherford, the organization, now called the Council for At-Risk Academics (CARA), enabled many notable twentieth-century scientists — including biochemist Hans Krebs and philosopher Karl Popper — escape the Nazis and settle at British universities. In recent years it has reached out to the Middle East and receives the largest volume of applications from Yemen and Iraq.

CARA and its counterparts in other countries exist because governments in the host nations value three of the pillars on which democracy rests: the rule of law, a free press and, as we explore in a Comment article on page 621, freedom of academic enquiry. If the British government were to decide not to support even one of these, CARA would struggle to carry on.

Such an alarming scenario is not purely hypothetical. For at least the past two decades, citizens of countries in the European Union have increasingly been voting for parties of the extreme right (also known as the populist right or radical right). From almost no representation in the 1990s, these parties are in governing coalitions in 10 out of the EU's 28 member states, including in Austria, Hungary, Italy and Poland. Next May sees elections to the European Parliament in which right-wing parties are expected to increase their combined tally of 78 seats in the 751 seat chamber.

When parties of either the extreme right or extreme left take power, any one of democracy's foundational pillars can be knocked away.

Journalists and their families are intimidated. Judges are demonized and replaced with allies. People from minority groups are singled out for their alleged disloyalty. And action is taken against academics: universities are brought under direct state control and staff are subjected to loyalty tests.

It's a classic playbook to quash dissent. Take Poland for example, where the state has moved to exert control over the media and judiciary. Academic freedom is under threat too. A barometer for the risk it could face will be how much protest the Polish government allows, if any, over its pro-coal stance — which climate scientists have warned against — during the annual United Nations climate talks to be held in Poland next month.

Although there has been much media attention on the phenomenon of the populist right, the implications for academic freedom have gone largely unreported. Even where there has been widespread coverage — such as the case of Hungary's Central European University which was forced to enrol new students in Vienna rather than Budapest — EU institutions such as the European Council and the European Parliament have been largely powerless to take action.

Europe's heads of government are biting their lips, and their reasons for doing so are understandable, even if European agreements or conventions are being violated. There is, of course, the principle of non-interference in the affairs of a sovereign state. But, in addition, the EU works through the collective solidarity of its member states. This is what has enabled the organization to enact progressive policies in climate change, anti-discrimination legislation and employee rights.

But collective progressivism breaks down when one-third of EU governments include political parties with scant commitment to protecting democratic institutions. Now that EU governments include parties who do not believe in the rights of people from minority groups, the consensus on climate change, or, indeed, academic freedom, it will become more difficult for the EU as a whole to either advance, advocate or protect policies in these fields.

> *"The implications for academic freedom have gone largely unreported."*

"What's wrong with the world is not nationalism itself," noted Michael Ignatieff, the embattled rector of the Central European University. What's wrong, he added, "is the kind of nation, the kind of home that nationalists want to create and the means they use to seek their ends."

Ignatieff wrote these words more than 20 years ago in *Blood and Belonging* (BBC Books, 1993), at the end of a series of journeys into some of Europe's conflict zones. But he remains optimistic about the continent's future. "I don't want to predict doom and gloom," he told *Nature*. "Regimes come and go, but universities remain."

Academics everywhere will hope he's right. They, and us, can help by speaking out against injustice and specific cases where academic freedom is threatened — by any regime. ■

Breeze block

Wind farms must be built responsibly so they don't create an inefficient wake for neighbours.

Like many words in the English language, 'overbearing' has a nautical connection. It describes a manoeuvre in which one sailing ship steers directly downwind towards another, effectively snatching away the overborne vessel's wind to leave it powerless.

Wind turbines can overbear each other, too. As developers seek to build ever more of them — globally, installed onshore wind capacity rose to almost 500 gigawatts last year, up from just 92 GW in 2007 — some of the best blustery locations are getting crowded. That could be a problem. To work best, wind turbines need to capture a clear and uninterrupted stream of moving air. Anything that gets in the way — from mountains and buildings to a rival wind farm — reduces wind speed and the electricity generated. Such obstacles also break up the

air flow and the resulting turbulence increases noise, as well as wear and tear on the turbine blades.

A study published in *Nature Energy* this week shows just how overbearing this effect can be (J. K. Lundquist *et al. Nature Energ.* https://doi.org/10.1038/s41560-018-0281-2; 2018). It analysed the change in electricity production at a wind farm in West Texas when another farm was built a few hundred metres upwind and switched on 18 months after the first farm opened. The researchers estimate that the downwind farm may have lost 5% of its potential on average, and as much as US$2 million annually in electricity production. Texas is unusual: it has the largest number of wind turbines in the United States, with more than 12,000 devices spread across 131 separate farms. Inevitably, the separate projects are clustering at the best sites, which have reliable wind and access to transmission lines. In the study, some of the turbines in the upwind farm stand just 300 metres from some of the downwind turbines.

But the study authors say the impact could stretch much further. Under the right atmospheric conditions, the decreases in downwind wind speeds can extend for 50 kilometres or more. Almost 90% of US wind farms have a neighbouring project within 40 km, and so could be affected. (Of course, not all of them would be affected all the time, because the wind changes direction. The Texas study looked only at the impact under the prevailing southwesterlies.) There is also inevitably internal disruption within a single wind project, with the upwind turbines creating a wake that reduces the output of those behind.

One solution to wind farms treading on each other's toes is to leave the land behind and head to the vast spaces of the oceans. But offshore wind farms — typically much more expensive to build and run — also tend to compete for the best sites. In 2014, the Danish firm DONG Energy Wind Power (now Ørsted, based in Skærbæk) published data to show how the performance of its long-standing project at Nysted, close to the island of Lolland in the Baltic Sea, was being undermined by a another company's wind farm constructed just 3 km away (N. G. Nygaard *J. Phys. Conf. Ser.* **524**, 012162; 2014).

What can be done? Technical fixes to the design or layout of projects are difficult, especially as wind turbines grow larger and more powerful. Some engineers have proposed offshore turbines that float and can shift position to reduce wake as the wind moves, but that's clearly impossible on land. Could rules and restrictions work? A legal analysis by the study authors found no relevant legislation in place in the United States. As a comparison, solar-power efficiency in California is protected by regulations to limit the amount of shadow from neighbouring properties that can fall on panels during peak operating hours.

> "It's crucial in a warming world to support efforts to boost wind power."

Where they exist, restrictions on the construction of wind turbines often focus on more immediate risks. In a 2008 dispute between rival developers who wanted to build wind farms on adjoining properties in North Dakota, officials ruled only that each turbine must be placed further than its own height from the boundary, so that if it fell it would not land on the other side. Wind shadow wasn't considered.

It's crucial in a warming world to support efforts to boost wind power, and therefore important to install wind farms responsibly to ensure that we harness as much energy as possible, even if the facilities are close together. That means it's important to craft regulations to support such development.

One country has long taken an enlightened view, and could offer a model to follow. The Netherlands is famous for its windmills, many of which still function, thanks to a law that guarantees each mill can continue to fill its sails with the necessary wind (called its *molenbiotoop*, or windmill biotope) by restricting development within 375 m. The law has led to some creative solutions: in 2010, a flour mill in Spijkenisse from the 1860s was cut from the ground, raised and placed on a 7 m-high concrete collar to allow houses to be built nearby. Where there's a mill, there's a way. ∎

Ban bullying

All institutions need a procedure for dealing with bullies.

Science can be difficult enough even if you work in a great laboratory with supportive colleagues. So the added pressure of a boss or co-worker who regularly abuses, trivializes, hassles, belittles and unfairly criticizes is not just a problem for the individual concerned. It's bad for research.

Such workplace bullying thrives on silence. But, as occurred with sexual harassment, there is growing noise about bullying in science. Already this year, allegations of bullying have rocked the world of astrophysics, closely followed by those of cancer genetics, neuroscience and vertebrate palaeontology.

Much of this additional scrutiny is down to the willingness of scientists to speak out. Now is the time for more institutions to follow their lead and step up to take decisive action. Does your institution have an anti-bullying policy? If you work in Britain, the answer is probably yes; but if you work in countries such as the United States, the answer might be no. As a News Feature on bullying in science highlights this week (page 616), few US institutions have policies that explicitly prohibit their staff from bullying others. Such behaviour might be covered by anti-harassment policies, but in those cases, targeted staff members can seek redress from their employer only if they fall into a group protected by employment law and can show that they have been targeted because of their sex, race, religion or age. The motivation of a bully should not be the issue here. Bullying is unacceptable, and more employers must make that clear.

What to do? If you feel that you are being singled out for unfair treatment by your boss or colleague, you have several options, and one of them is to talk to others. You will need support from your friends and family, and no one can help you if they don't know it's happening. By sharing your story with trusted peers, you might discover that other people you work with are going through the same thing.

Seek advice about what you can do to address the problem. Speak to someone in your institution's human-resources department or a manager about how to solve the problem informally. If you belong to a union, you can ask it for advice. It can be helpful to keep a diary of the problematic behaviour. If you feel confident enough and it is safe to do so, think about speaking to the bully. Calmly try to tell them that you find their behaviour unacceptable and ask them to stop.

Many who have been through the process can testify to the professional upheaval and emotional turmoil that comes with reporting a bully. It is easy for those who are not in sitting in the eye of the storm to extol the virtues of flagging up bullying for the greater good of science and society. There are no easy answers, and some cases might boil down to one person's word against another's.

This is why institutions need to step up to the mark. Reports of bullying should be fairly and thoroughly investigated, with attention to due process. Anti-bullying policies or codes of conduct for staff should be easily accessible, give clear guidance on what behaviours are and are not appropriate in the workplace, and outline the measures that would be taken if allegations are reported.

Crucially, institutions need to follow these policies to the letter, regardless of whether the alleged perpetrator is the director of the institute or a first-year PhD student, to protect all those involved — including the accused, who might be the victim of malicious allegations. Incomplete or unfair investigations can undermine the credibility of an organization, harm careers and signal to bullies that their behaviour will be tolerated — in 2018 that is unacceptable. ∎

WORLD VIEW
A personal take on events

Be human first, a scientist second

Want to get the best research from your team? Take these six steps to invest in stronger relationships, urges **Alison Antes**.

"I was a human first, and then I learned to be a scientist. If I forget the human part, then that's a problem."

This is what I heard when I interviewed 52 scientists recognized as exemplary by their peers for their scientific accomplishments and conduct. Related themes come up in my work with scientists who have been referred to a formal remediation programme after lapses in research integrity.

I'm an organizational psychologist, specializing in the scientific workplace. What interests me are the decisions and behaviours that yield innovative, rigorous, ethical research.

The past few months have drawn attention to unhealthy working environments, especially bullying in academia. We should also focus on a related, widespread problem: mentors who have excellent intentions but limited knowledge of how to create a healthy workplace.

Many scientists whom I work with feel that they lack management and leadership skills. They want help with concrete tasks such as coordinating projects or facilitating meetings. But what comes up most emphatically is that conducting research requires them to establish and maintain positive relationships in the lab.

Many researchers in our remediation programme have had strained interactions with compliance officers and have struggled in their roles as supervisors. By contrast, exemplars resoundingly emphasize how they foster good team dynamics by being involved, approachable and aware of the workplace atmosphere. As one told me: "Rule number one in the lab is harmony. First and foremost, we have to get along, we have to respect each other, we have to trust each other, and that is the operating principle for everything else."

Yet, given the choice between working on a scientific paper or broaching a difficult conversation, many researchers pick the former — the task that feels most directly connected to research goals. Principal investigators might need to work consciously against the feeling that 'nothing is getting done' during personal interactions. Because, whether it is mentoring a struggling trainee or celebrating a hard-won achievement, investing in strong, respectful relationships is an investment in effective science.

So, what to do? All principal investigators should add relationship building to their to-do lists.

Task one: put recurring one-on-one meetings with the members of your group on your calendar. Set up a notebook or spreadsheet and jot down anything you should bring up during these meetings. Set an alert for ten minutes before the appointment to decide how to approach the meeting. Does the team member need encouragement? Career guidance? Feedback on their project and direction for next steps? Are they behind on deadlines or lacking confidence? Try a respectful, yet firm, nudge. Have you expressed gratitude for their contribution? As

one exemplar noted: "I value what they do, and I tell them."

Ask yourself whether it is time for a difficult conversation. If so, grasp the nettle. That is part of a leader's job. Sometimes principal investigators worry that they will damage relationships by having challenging discussions. In the long-run the opposite is true. Use your ten minutes to list a few observations. State the specific behaviour of concern; describe how it affected you, the team or the project. Then, ask the person for their perspective. If there is discord in the lab, speak to the individuals involved, state your expectation of mutual respect, ask them to discuss and identify a solution.

Task two: invite people to share both complaints and highlights. Several exemplary scientists explicitly require their trainees to relate a concern or struggle at some point in one-on-one meetings. They want to help people to be comfortable enough to bring problems and mistakes to light, and so address issues early, while they are manageable. Several exemplars noted that researchers need outlets for discussing frustrations and anxieties. They know it is difficult to show up and do your best when plagued by worry. And they want to know what is working well in the lab, so as to leverage these successes.

> SHOWING THAT **YOU CARE** IS MORE IMPORTANT THAN SHOWING THAT YOU ARE **PERFECT.**

Task three: walk the 'shop floor'. Even when team members are welcome to visit your office, visibility supports approachability, impromptu brainstorming and immediate trouble-shooting.

Task four: model desired behaviour in team meetings. How you communicate will carry over into peer-to-peer interaction in your group. Ask questions, expect participation and prompt people to share their thoughts. Find out where obstacles are. Encourage cooperation and mutual support. Explicitly state that you value a collaborative spirit in your group.

Task five: schedule a few social occasions for people to spend time together in a more relaxed way. Such activities might feel far removed from science, but they can ease tensions in the lab. Start small. Be sure to accommodate the needs of parents and carers, people with cultural or religious considerations and those on tight budgets.

Task six: advocate outside the lab. Talk about these practices in your department, share those that work and learn from people known to be great team leaders.

New principal investigators commonly adopt the practices of their own mentors without reflection, and often their role models were not ideal. Some relationship-building tasks will feel awkward at first; that's okay. Showing that you care is more important than showing that you are perfect. ∎

Alison Antes is assistant director of the Center for Clinical and Research Ethics at Washington University School of Medicine in St. Louis, Missouri.
e-mail: aantes@wustl.edu

SEVEN DAYS

The news in brief

Scotland's research

The Scottish government and seven bodies representing the nation's research and higher-education sectors — including Universities Scotland and the Royal Society of Edinburgh — have agreed to work together to protect Scottish research from Brexit. The alliance released a joint statement on 22 November, coinciding with a summit at the University of Glasgow that attracted research leaders from across Scotland. The group aims to use its influence to press the UK government for clarifications and firmer guarantees on research, as well as to give Scotland a visa system that allows overseas students and postgraduates to stay on and work after their studies. Around 27% of its full-time researchers and 10% of its university students are nationals of other European Union countries. Once Britain leaves the EU, it is likely to be harder for them to come to the United Kingdom to work.

Brexit framework

European Union leaders meeting in Brussels on 25 November approved a declaration on the EU's future relationship with the United Kingdom. Billed as a framework that will form the basis of a trade deal beyond the end of a transition period in December 2020, the 26-page document confirms Britain's intention to end free movement across its borders. It includes a pledge by both parties to consider arrangements "for entry and stay for purposes such as research, study, training and youth exchanges". The document reiterates the United Kingdom's intention to pay to participate in EU programmes in areas including science and innovation. And it says that the country aims to remain a

Dire US climate-change forecast

Climate change is already affecting life in the United States, and its impacts are set to become more dramatic in the coming decades, according to the US government's fourth national climate assessment. The analysis, produced by 13 federal agencies and required by law every 4 years, was released on 23 November. Among other things, the report finds that higher temperatures and drier conditions have led to more large fires in the western United States (pictured), and the combination of rising seas and extreme precipitation has boosted flood risks along the east coast. The core message of the document contradicts positions taken by US President Donald Trump's administration, which released the report on Black Friday, the day after the US Thanksgiving holiday. Some scientists and environmentalists have suggested that this timing was part of an attempt to downplay the report's findings.

part of the European Research Infrastructure Consortium of research networks, two of which it currently hosts, and of the European Defence Fund, a scheme established last year whose research budget could rise to around €500 million (US$570 million) a year from 2021. At the same meeting, the EU leaders endorsed a withdrawal agreement, published on 14 November, which lays out the terms of Britain's exit from the EU.

China's mega 'LHC'

Beijing's Institute of High Energy Physics (IHEP) is designing the world's biggest particle smasher. The 100-kilometre-circumference facility would dwarf the 27-kilometre Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. The ambitious 30-billion-yuan (US$4.3-billion) facility, known as the Circular Electron–Positron Collider, is the brainchild of IHEP's director, Wang Yifang. The collider will produce Higgs bosons by smashing together electrons and their antimatter counterparts, positrons. Because these are fundamental particles, their collisions are cleaner and easier to decipher than the LHC's proton–proton collisions, so once the Chinese facility opens, in about 2030, it will allow physicists to study the mysterious particle and its decay in exquisite detail. Initial funding has come from the Chinese government, but the design is the work of an international collaboration of physicists that hopes to garner international funding. The blueprints published on 14 November reveal that the collider would run in a circle 100 metres underground, at a location yet to be decided, and would host two detectors.

Science in a cloud

A preliminary version of the European Open Science Cloud online portal launched

on 23 November. Scheduled to become available in full in 2020, the portal is intended to make it easier for European researchers to store, analyse, share and reuse data. The launch of www.eosc-portal.eu — which will eventually provide a single entry point to existing data repositories, as well as cloud-computing facilities and analysis tools — comes after two years of consultation and development. At the launch in Vienna, the European Commission also announced the make-up of the initiative's executive board, which will comprise representatives of university associations, data infrastructures and research institutes, as well as three independent experts. The commission plans to allocate €600 million (US$680 million) to the initiative by 2020.

African artefacts

Tens of thousands of African artefacts in French museums should be returned, concludes a 23 November report commissioned by the country's president, Emmanuel Macron. The report — by economist Felwine Sarr at Gaston Berger University in Saint-Louis, Senegal, and historian Bénédicte Savoy of the College de France in Paris — calls on France to amend its laws to allow for the repatriation of cultural artefacts acquired during the French colonial period in Africa (**pictured**), if African countries request their return. This includes artefacts from the late nineteenth century until 1960 and those later acquired illicitly. The Quai Branly Museum in Paris holds some 70,000 objects from sub-Saharan Africa.

Salk settlement

The prestigious Salk Institute for Biological Studies in La Jolla, California, has settled the final one of three high-profile gender-discrimination lawsuits filed last year. The agreement was announced on 21 November. Molecular biologist Beverly Emerson filed the suit in July 2017, arguing that discrimination against women at the Salk had limited her wages, laboratory space and research funding. Two other senior female scientists brought similar suits against the institute, and settled their cases out of court in August 2018. Emerson had worked at the Salk for more than three decades, but in December last year, the institute declined to renew her contract. She is now at Oregon Health and Science University in Portland. "Salk recognizes Dr. Emerson's more than thirty years of service to the Institute and looks forward to her continued contributions to the scientific community," says a joint statement from Emerson and the institute, e-mailed to *Nature*. The statement does not provide any further information on the settlement. Alreen Haeggquist, Emerson's lawyer, says that neither she nor Emerson has further comment.

Plan S detailed

A group of 16 science funders have detailed their ambitious plan to ensure that, by 2020, the results of the research they support is immediately free to read. Since the September launch of the initiative, known as Plan S, scientists have speculated as to how it could affect their research. Many publishers have also expressed serious concerns about the proposal and have questioned its rationale for excluding 'hybrid journals' — journals that allow researchers to make their work free to read if they pay a fee, but that keep other studies behind a paywall. Now, documents released on 26 November clarify that researchers will be allowed to publish in hybrid journals if they can post the accepted manuscript or final article in an approved open-access repository at the time of publication — but in these cases, the funder will not pay for publishing. The plan's documents also list three ways in which researchers can publish work that is compliant with the plan: publish in an open-access journal or platform approved by the funders; immediately put a copy of the manuscript accepted by the journal, or the final published article, in an approved open-access repository; or use a hybrid journal that intends to become a full open-access venue. Under all three routes, the papers must be published with a liberal CC BY licence, which allows commercial reuse of the papers' findings. The posting of an article to a preprint server is not, alone, sufficient to comply with these rules.

TREND WATCH

Biomedical research is becoming more open and transparent by providing increasing amounts of information about funding, conflicts of interest and data sharing in its publications, according to a survey of recent papers. John Ioannidis at Stanford University in California and colleagues examined 149 papers published between 2015 and 2017 to see how many included information on indicators of transparency, such as who funded the work, potential conflicts of interest, and the availability of the raw data and complete research protocols.

They found that the majority of papers contained statements on funding and conflicts of interest (69% and 65%, respectively), and almost one in five mentioned publicly available data — although only one paper included a link to a full study protocol (J. D. Wallach *et al. PLoS Biol.* http://doi.org/cxd6; 2018).

The survey results were a big improvement on those of a previous study by some of the same researchers. This found that, in a sample of 441 articles published between 2000 and 2014, most contained almost no information on funding, conflicts of interest or data sharing.

OPENING UP

The proportion of biomedical journal articles that provide information on their funding, conflicts of interests and data is on the rise.
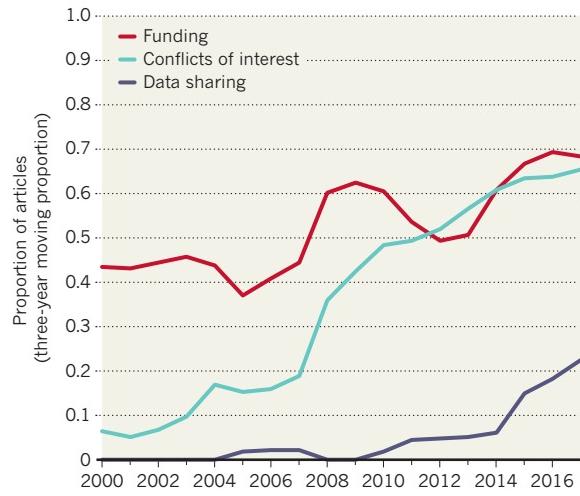

Legend:
- Funding
- Conflicts of interest
- Data sharing

Y-axis: Proportion of articles (three-year moving proportion), 0 to 1.0
X-axis: 2000 to 2016

NEWS IN FOCUS

PASCAL GOETGHELUCK/SPL

A Chinese scientist claims that twin girls have been born whose genomes were edited at the embryo stage.

GENOMICS

International outcry over genome–edited baby claim

The revelation from a Chinese scientist represents a controversial leap in genome editing.

BY DAVID CYRANOSKI & HEIDI LEDFORD

Scientists are shocked and outraged by reports that a Chinese scientist claims to have helped make the world's first genome-edited babies — twin girls, who were born this month.

He Jiankui, a genome-editing researcher at the Southern University of Science and Technology of China in Shenzhen, says that he impregnated a woman with embryos that had been edited to disable the genetic pathway HIV uses to infect cells.

In a video posted to YouTube on 26 November, He says the girls are healthy and now at home with their parents. Sequencing of the babies' DNA has shown that the editing worked, and altered only the target gene, he says. The scientist's claims have not been verified through independent genome testing, nor published in a peer-reviewed journal. Later that day, the Chinese government announced an investigation into the claims.

If the report is true, the twins' birth would represent a significant — and controversial — leap in the use of genome editing. Until now,

the use of these tools in embryos has been limited to research, often to investigate the benefit of using the technology to eliminate disease-causing mutations from the human germ line. But some studies have reported off-target effects, raising significant safety concerns.

Documents posted on China's clinical-trial registry show that He used the popular CRISPR–Cas9 genome-editing tool to disable a gene called *CCR5*, which encodes a protein that allows HIV to enter a cell. Genome-editing scientist Fyodor Urnov was asked to review documents that described the ▶

claimed experiments for an article in *MIT Technology Review*. "The data I reviewed are consistent with the fact that the editing has, in fact, taken place," says Urnov, who is based at the Altius Institute for Biomedical Sciences in Seattle, Washington. But he adds that the only way to tell whether the children's genomes have been edited is to independently test their DNA.

Urnov takes issue with the decision to edit an embryo's genome to prevent HIV infection. He is also using genome-editing tools to target the *CCR5* gene, but his studies are in people with HIV, not embryos. He says that there are "safe and effective ways" to use genetics to protect people from HIV that do not involve editing an embryo's genes.

Paula Cannon, who studies HIV at the University of Southern California in Los Angeles, also questions He's decision to target that gene in embryos. She says that some strains of HIV don't even use this protein to enter cells, they use another protein called CXCR4. Even people who are naturally CCR5-negative are not completely resistant to HIV, Cannon adds, because they could be infected by a CXCR4 strain.

She also says it makes no sense that He recruited families with an HIV-positive father, as was the case with the twins, because there is no real risk of transmission to the children.

"This experiment exposes healthy normal children to risks of gene editing for no real necessary benefit," says Julian Savulescu, director of the Oxford Uehiro Centre for Practical Ethics at the University of Oxford, UK.

In an interview with the Associated Press, He said the goal of the work was not to prevent transmission from the parents, but to offer couples affected by HIV a chance to have a child that might be protected from a similar fate. But years of research is needed to show that meddling with the genome of an embryo is not going to cause harm, says Joyce Harper, who studies women's and reproductive health at University College London. Legislation and public discussion should also occur before genome editing is used in embryos destined for implantation.

Southern University of Science and Technology said in a statement on 26 November that it was unaware of He's experiments, that the work was not performed at the university and that He has been on leave since February. The university says its researchers must abide by national laws and regulations, and respect international academic ethics and academic standards. It will set up an independent committee to investigate the matter.

Making gene-edited babies goes against regulations released by China's health and science ministries in 2003, but it is not clear whether there are penalties for those who break the rules.

More than 100 Chinese biomedical researchers posted a strongly worded statement

> *"This is a huge blow to the international reputation and the development of Chinese science."*

online condemning He's claims. "Directly jumping into human experiments can only be described as crazy," the statement reads. The scientists call on Chinese authorities to release the findings of any investigation to the public.

"This is a huge blow to the international reputation and the development of Chinese science, especially in the field of biomedical research," the statement says. "It is extremely unfair to the large majority of diligent and conscientious scientists in China who are pursuing research and innovation while strictly adhering to ethical limits."

Nature tried to contact He but did not receive a response before its deadline. In his video, He says he supports the use of genome editing in embryos only in cases that relate to disease. "I understand my work will be controversial, but I believe families need this technology and I am willing to take the criticism for them," he says.

News of the experiment came a day before researchers in the field gathered in Hong Kong for a major international meeting on genome editing, running from 27 to 29 November. Even before the news of He's work emerged, many in the field thought it was inevitable that someone would use genome-editing tools to make changes to human embryos for implantation into women, and had been pushing for an international consensus on how genome editing to modify eggs, sperm or embryos should proceed. ■

'Marsquake' hunter begins to probe planet's innards

Joint US–French–German mission will monitor seismic activity on Mars.

BY ALEXANDRA WITZE

Earthlings are about to hear Mars's heartbeat.

On 26 November, NASA's InSight mission touched down near the Martian equator and embarked on the first mission dedicated to listening for seismic energy rippling through the red planet.

Any 'marsquakes' InSight detects could yield clues about the planet's mysterious interior, including how it is separated into a core, mantle and crust. Whatever scientists learn about Mars's innards could help to illuminate how our own planet evolved billions of years ago.

InSight had been cruising through space since its launch in May, tracked by mission control at NASA's Jet Propulsion Laboratory

(JPL) in Pasadena, California. On Monday, just before 11:53 a.m. local time, the spacecraft entered the Martian atmosphere at nearly 20,000 kilometres per hour.

As it neared Mars's surface, the spacecraft demonstrated a new way to communicate with its controllers on Earth, 146 million kilometres away. Two 'cubesats', each the size of a briefcase, relayed information from InSight to Earth in close to real time. The experiment suggests that miniature satellites like these could allow faster communication with probes in deep space.

InSight landed at Elysium Planitia, a broad, flat region just north of the Martian equator. It is one of the most boring places on the planet, says Bruce Banerdt, a planetary scientist at JPL and the US$994-million mission's principal investigator. That's an advantage for InSight,

which needs a safe, geologically stable place to do its work.

The first photo that InSight sent from the surface of Mars showed a flat, relatively rock-free landscape stretching to the horizon, with the foreground speckled with dust from the landing.

"It's happy. The lander is not complaining," said Rob Manning, chief engineer at JPL, shortly after InSight touched down.

LISTENING IN

Mission scientists will use the lander's camera to scout the ground for the smoothest and most level area to deploy its French-built seismometer (see 'Ear to the ground'). InSight's robotic arm will pluck the instrument off its back and place it on the ground, then put a

EAR TO THE GROUND
NASA's Mars InSight lander will gather data on seismic activity to help scientists better understand the red planet's mysterious interior.

The 1.8-metre robotic arm will place a seismometer and heat probe onto Mars's surface.

Two solar panels will supply power to the lander and its instruments.

The lander's seismometer will listen for tremors known as marsquakes.

A heat probe will dig down 5 metres to measure temperature change over depth and time.
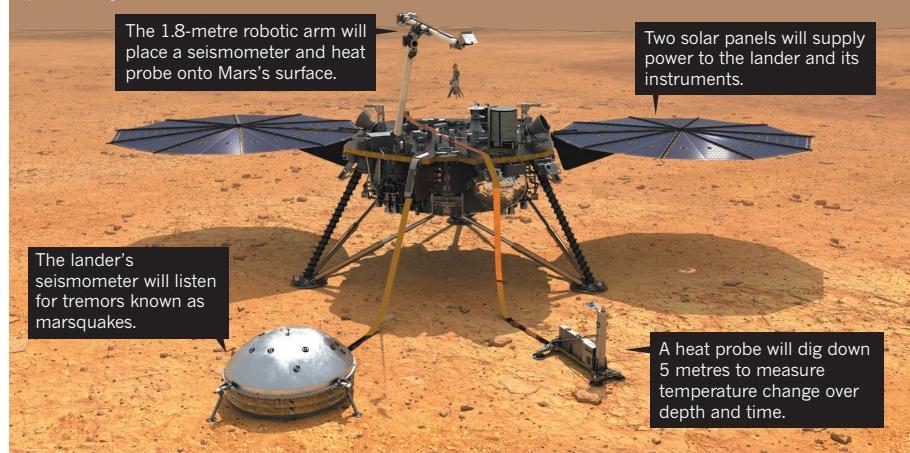
dome-shaped wind shield over it. The whole process is expected to take several days.

The seismometer includes three ground-motion sensors nested inside a vacuum, and its sensitivity allows it to detect movement as small as the width of an atom. The big challenge will be determining which movements are caused by marsquakes and which are the result of jostling by the wind or other sources. On the third day after landing, project scientists will switch on an instrument to track changes in the magnetic field, which will help them to identify sources of noise that aren't quakes, says Catherine Johnson, a geophysicist at the University of British Columbia in Vancouver, Canada.

InSight won't deploy its German-built heat-flow probe until January. Over the course of several weeks, the instrument will drill five metres into the Martian surface, deeper than anything achieved before. Scientists will track changes in temperature as small as a few hundredths of a degree. That will tell them how much heat is leaving Mars, and how many heat-producing radioactive elements are packed inside it.

InSight is meant to work for a little more than one Martian year, equivalent to almost two Earth years. It should measure 50–100 marsquakes during that period, says Banerdt. The longer it survives, the more it will be able to detect — and the more researchers will be able to deduce about Mars's internal structure. ∎

PUBLISHING

The age of AI peer reviews

Automated software can help review papers, but the decision–making stays with humans.

BY DOUGLAS HEAVEN

Most researchers have good reason to grumble about peer review: it is time-consuming and error-prone, and the workload is unevenly spread, with just 20% of scientists taking on most reviews.

Now peer review by artificial intelligence (AI) is promising to improve the process, boost the quality of published papers — and save reviewers time. A handful of academic publishers are piloting AI tools to do anything from selecting reviewers to checking statistics and summarizing a paper's findings.

In June, software called StatReviewer, which checks that statistics and methods in manuscripts are sound, was adopted by Aries Systems, a peer-review management system owned by Amsterdam-based publishing giant Elsevier. And ScholarOne, a peer-review platform used by many journals, is teaming up with UNSILO of Aarhus, Denmark, which uses natural language processing and machine learning to analyse manuscripts.

UNSILO uses semantic analysis of the manuscript text to extract what it identifies as the main statements. This gives a better overview of a paper than the keywords typically submitted by authors, says Neil Christensen, sales director at UNSILO. "We find the important phrases in what they have actually written," he says, "instead of just taking what they've come up with five minutes before submission."

UNSILO identifies which of these key phrases are most likely to be claims or findings, giving editors an at-a-glance summary of the results. It also highlights whether the claims are similar to those from previous papers, which could be used to detect plagiarism or simply to place the manuscript in context with related work in the wider literature. "The tool's not making a decision," says Christensen. "It's just saying: 'Here are some things that stand out when comparing this manuscript with everything that's been published before. You be the judge.'"

"It doesn't replace editorial judgement but, by God, it makes it easier," says David Worlock, a UK-based publishing consultant who saw the UNSILO demonstration at the Frankfurt Book Fair in Germany last month.

Worlock notes that there are several similar tools emerging. He is on the board of Wizdom.ai in London, a start-up owned by publishers Taylor & Francis, which is developing software that can mine paper databases and extract connections between different disciplines and concepts. He says that this kind of tool will be useful beyond peer review, for tasks such as writing grant applications or literature reviews.

Many platforms, including ScholarOne, already have automatic plagiarism checkers. And services including Penelope.ai examine whether the references and the structure of a manuscript meet a journal's requirements. Some can flag up issues with the quality of a study, too. The tool statcheck, developed by Michèle Nuijten, a methodologist at Tilburg University in the Netherlands, and her colleagues, assesses the consistency of authors' statistics reporting, focusing on *P* values. The journal *Psychological Science* runs all its papers through the tool, and Nuijten says that other publishers are keen to integrate it into their review processes.

When Nuijten's team analysed papers published in psychology journals, they found that roughly 50% contained at least one statistical inconsistency (M. B. Nuijten *et al. Behav. Res. Meth.* **48**, 1205–1226; 2016). In one in eight papers, the error was serious enough that it could have changed the statistical significance of a published result. "That's worrisome," she says. She's not surprised that reviewers miss such mistakes, however. "Not everyone has time to go over all the numbers. You focus on the main findings or the general story."

For now, statcheck is limited to analysing ▶

"It doesn't replace editorial judgement but, by God, it makes it easier."

Automation of standardized tasks could take the slog out of peer review.

manuscripts that use the American Psychological Association's reporting style for statistics. By contrast, the creators of StatReviewer, Timothy Houle at Wake Forest University School of Medicine in North Carolina and Chadwick DeVoss, chief executive of tech start-up NEX7 in Madison, Wisconsin, say that their tool can assess statistics in standard formats and presentation styles from multiple fields. To do this, it checks that papers correctly include things such as sample sizes, information about blinding of experiments and baseline data.

DETECTING FRAUD MARKERS
StatReviewer can also identify markers of fraudulent behaviour, says DeVoss. "Things like, did they game some statistical rules, or did they flat-out make up data? If the risk is higher than what the journal is used to seeing, they can look into the details." DeVoss says that StatReviewer is being tested by dozens of publishers. A 2017 trial with the open-access publisher BioMed Central in London was inconclusive because the tool did not analyse enough manuscripts, but did nonetheless provide some insights. BioMed Central is now planning a follow-up.

StatReviewer did catch things that human reviewers missed, says Amy Bourke-Waite, communications director for open research at Springer Nature, which owns BioMed Central and publishes *Nature* (*Nature*'s news team is editorially independent of Springer Nature). For example, it was good at catching papers that did not meet required standards, such as following CONSORT, a manuscript format used by many publishers. Bourke-Waite adds that authors who took part said that they were as happy responding to StatReviewer reports as they were to the human reviewer's. Occasionally, she says, StatReviewer got things wrong — but sometimes its slip-ups drew authors' attention to unclear reporting in their manuscripts.

Even if the trials prove successful, DeVoss expects that only some journals will want to pay to have all their manuscripts scanned. So he and his colleagues are targeting authors, too, hoping that they will use the tool to check their manuscripts before submission.

There are potential pitfalls to AI in peer review in general. One concern is that machine-learning tools trained on previously published papers could reinforce existing biases in peer review. "If you build a decision-making system based on the articles which your journal has accepted in the past, it will have in-built biases," says Worlock. And if an algorithm provides a single overall score after evaluating a paper, as StatReviewer does, there might be a temptation for editors to cut corners and simply rely on that score in deciding to reject a paper, says DeVoss.

Algorithms are not yet smart enough to allow an editor to accept or reject a paper solely on the basis of the information they extract, says Andrew Preston, co-founder of Publons, a Wellington-based start-up acquired by Clarivate Analytics in Philadelphia, Pennsylvania, that tracks peer review and is using machine learning to develop a tool to recommend reviewers. "These tools can make sure a manuscript is up to scratch, but in no way are they replacing what a reviewer would do in terms of evaluation."

Nuijten agrees: "The algorithms are going to need some time to perfect, but it makes sense to automate a lot of things, because a lot of things in peer review are standard." ∎

Can conference shed reputation for hosting sexist behaviour?

AI meeting wants to become more inclusive, but survey suggests it has a long way to go.

BY HOLLY ELSE

Hordes of artificial-intelligence researchers will descend this weekend on one of the field's hottest tickets: the Neural Information Processing Systems conference in Montreal, Canada. But although attendees at this annual event will hear talks on cutting-edge ideas in computer science, another issue will also be front and centre: whether the conference can provide a welcoming environment for women as the field of artificial intelligence (AI) grapples with a culture of harassment and discrimination.

The concerns were thrown into stark relief earlier this month with the release of a survey of 2,375 people — most of whom had either attended the meeting or submitted papers for consideration in previous years.

Respondents reported experiencing sexual harassment, seeing the conference welcome sexist people and regularly hearing sexist or sexually abusive comments and jokes. Women

reported unwelcome, persistent advances from men at the conference. The analysis does not reveal what percentages of respondents reported these experiences, but does say that 15% of respondents were women.

Terrence Sejnowski, president of the foundation that oversees the conference, told *Nature* that the foundation's board, and others, had read the report with great interest, and thanked the authors for the analysis. "It provides us with valuable information for understanding our community," he said.

DIVERSITY MEASURES
The survey was carried out by Katherine Heller, a machine-learning researcher at Duke University in Durham, North Carolina, and Hal Daumé, a machine-learning researcher at the University of Maryland in College Park, who are the diversity and inclusion chairs at this year's event.

In December 2017, Sejnowski and the chairs of the boards of the 2017 and 2018 conferences acknowledged that several events held at or in conjunction with the 2017 conference had fallen short of the standards required to "provide an inclusive and welcoming environment for everyone". They said that they would take immediate action, including recruiting the

diversity and inclusion chairs, formalizing the process for reporting concerns and strengthening an existing code of conduct, by which all attendees and sponsors will have to abide in future.

Their statement came shortly after several female machine-learning researchers spoke out about their experiences at last year's event in Long Beach, California, and other AI conferences, including a joke about sexual assault, allegedly made by a member of a band composed of leading researchers at a party coinciding with the 2017 event.

Other measures to improve inclusion include subsidized childcare and a diversity meeting. There are also now several ways for conference-goers with concerns to notify organizers.

And on 16 November, the board abandoned the commonly used acronym, NIPS, and renamed the event NeurIPS. A March 2018 letter to the board, signed by 122 academics at Johns Hopkins University in Baltimore, Maryland, said the NIPS acronym was "prone to unwelcome puns" and revealed further goings-on at the conference, including an unofficial sister event named "TITS" and T-shirts spotted bearing the slogan "my NIPS are NP-hard".

Researchers have mixed views about

whether the board's efforts will bring meaningful change. Raia Hadsell, a machine-learning researcher at DeepMind in London who has been attending the conference for more than a decade has not witnessed a "rampant culture of discrimination, bias or harassment" at the event but has seen and experienced problematic behaviour. "I find it infuriating to be asked whether I am a recruiter, or a 'plus one', or whether I 'did the work myself' — do men ever, ever get asked questions like that?" she says.

She thinks that the machine-learning community wants to address the problems, but that their complexity makes it difficult. "I think that there will still be a problem come December in Montreal."

Elana Fertig, a computational biologist at Johns Hopkins University who signed the March letter to the board, says that altering the name is a powerful first step that has heightened awareness of the issues and shows that change is possible. But two of Fertig's students decided earlier this year not to attend the event because of the reported culture. And she worries about a backlash against the name change, noting that there were negative, sometimes threatening, comments that accompanied the debate over the change. ∎

Alzheimer's researchers seek better mice

Several teams are developing animal models that more closely mimic the disease in people.

BY SARA REARDON

Drug companies have spent billions of dollars searching for therapies to reverse or significantly slow Alzheimer's disease, to no avail. Some researchers argue that the best way to make progress is to create better animal models for research, and several teams are now developing mice that more closely simulate how the disease devastates people's brains.

The US National Institutes of Health (NIH), the UK Dementia Research Institute and the Jackson Laboratory — one of the world's biggest suppliers of laboratory mice — are among the groups trying to genetically engineer more-sophisticated rodents. Scientists are also probing the complex web of mutations that influence neurological decline in mice and people.

"We appreciate that the models we had were insufficient," says Bruce Lamb, a neuroscientist

at Indiana University in Indianapolis who directs the NIH-funded programme. "I think it's sort of at a critical juncture right now."

Alzheimer's is marked by cognitive decline and the build-up of amyloid-protein plaques in the brains of people, but the disease does not occur naturally in mice. Scientists get around this by studying mice that have been genetically modified to produce high levels of human amyloid protein. These mice develop brain plaques, but no memory problems.

"I think it's sort of at a critical juncture right now."

Many experimental drugs that have successfully removed plaques from mouse brains have not lessened the symptoms of Alzheimer's disease in people. One high-profile stumble came last month, when three companies reported that their Alzheimer's drugs — from a class called BACE inhibitors — had failed in late-stage

clinical trials. Although the drugs successfully blocked the accumulation of amyloid protein in mice, they seemed to worsen cognitive decline and brain shrinkage in people.

The drive for better mice comes as genomics studies are linking the most common form of Alzheimer's — late onset — to dozens of different genes. This diversity suggests that each case of the disease is caused by a different mix of genetic and environmental factors. "There is no single Alzheimer's disease," says Gareth Howell, a neuroscientist at the Jackson Laboratory in Bar Harbor, Maine.

Howell argues that scientists' reliance on inbred lab mice with only a few engineered mutations might have limited research. His own work suggests that, in mice, just as in people, genetic diversity plays a part in determining how neurodegeneration progresses.

When Howell's team modified two genes associated with early-onset Alzheimer's in both lab mice and their wild cousins, all of ▶

the animals developed amyloid plaques. But although the more-inbred lab mice did not display any outward signs of Alzheimer's, a portion of the genetically diverse wild mice experienced memory problems. The researchers think that a combination of plaques and unknown genetic factors caused these symptoms. They presented the results this month at a meeting of the Society for Neuroscience in San Diego, California.

Another study, by neuroscientist Catherine Kaczorowski at the Jackson Laboratory, suggests that animals' genetic make-up affects how they respond to environmental triggers. Her group bred genetically diverse wild mice with lab mice that had mutations that cause amyloid plaques to form. Some of the resulting offspring were more likely to develop cognitive problems if they ate a high-fat diet, but other mice on the diet had a lower risk of these symptoms, Kaczorowski reported at the San Diego meeting.

An Alzheimer's model mouse.

Understanding how this expanded universe of genetic factors affects Alzheimer's risk will require a host of new animal models with different combinations of mutations. Several efforts to engineer these next-generation mice are already under way.

In 2016, the NIH started the MODEL-AD consortium to develop more Alzheimer's mice and make them available to researchers. Project scientists engineer mice with different genetic mutations associated with early- or late-onset Alzheimer's, and test the animals to see whether they display signs of the disease. They then post descriptions of each mouse type in an online database. Lamb says that the team has released about 30 mouse varieties, and received more than 500 orders for the animals from academic scientists and biotechnology firms.

And in January, the UK Dementia Research Institute in London launched a similar programme. Scientists there are developing model mice whose brains show the amyloid plaques and tangles of another protein, called tau, that occur in people with Alzheimer's. To mimic the brain

inflammation that the disease causes, the group is implanting neural immune cells grown from human stem cells into the brains of mice.

Ultimately, researchers hope that the models will reveal ways to predict whether a person will respond to a particular Alzheimer's therapy. And having a better understanding of how inflammation and genes drive the disease could help to identify it in people before plaques and tangles have formed, says Rudolph Tanzi, a neurologist at Harvard University in Cambridge, Massachusetts. "That's why it's so important to have those animal models available and really start working on all these genes."

But Bart de Strooper, a molecular biologist at the Catholic University of Leuven in Belgium, urges caution. De Strooper, who directs the UK programme, says that none of the next-generation animals is likely to be a perfect analogue for people. "The biggest mistake you can make," he says, "is to think you can ever have a mouse with Alzheimer's disease." ■

JACKSON LABORATORY

CORRECTION

The News Feature 'Why extreme rains are getting worse' (*Nature* **563**, 458–460; 2018) erroneously located Elizabeth Kendon in Reading. She is, in fact, at the Met Office in Exeter.

THE SUN DIMMERS

*With dire climate scenarios on the horizon, researchers are getting serious about solar geoengineering.*

BY JEFF TOLLEFSON

Zhen Dai holds up a small glass tube coated with a white powder: calcium carbonate, a ubiquitous compound used in everything from paper and cement to toothpaste and cake mixes. Plop a tablet of it into water, and the result is a fizzy antacid that calms the stomach. The question for Dai, a doctoral candidate at Harvard University in Cambridge, Massachusetts, and her colleagues is whether this innocuous substance could also help humanity to relieve the ultimate case of indigestion: global warming caused by greenhouse-gas pollution.

The idea is simple: spray a bunch of particles into the stratosphere, and they will cool the planet by reflecting some of the Sun's rays back into space. Scientists have already witnessed the principle in action. When Mount Pinatubo erupted in the Philippines in 1991, it injected an estimated 20 million tonnes of sulfur dioxide into the stratosphere — the atmospheric layer that stretches from about 10 to 50 kilometres above Earth's surface. The eruption created a haze of sulfate particles that cooled the planet by around 0.5 °C. For about 18 months, Earth's average temperature returned to what it was before the arrival of the steam engine.

The idea that humans might turn down Earth's thermostat by similar, artificial means is several decades old. It fits into a broader class of planet-cooling schemes known as geoengineering that have long generated intense debate and, in some cases, fear.

Researchers have largely restricted their work on such tactics to computer models. Among the concerns is that dimming the Sun could backfire, or at least strongly disadvantage some areas of the world by, for example, robbing crops of sunlight and shifting rain patterns.

But as emissions continue to rise and climate projections remain dire, conversations about geoengineering research are starting to gain more traction among scientists, policymakers and some environmentalists. That's because many researchers have come to the alarming conclusion that the only way to prevent the severe impacts of global warming will be either to suck massive amounts of carbon dioxide out of the atmosphere or to cool the planet artificially. Or, perhaps more likely, both.

If all goes as planned, the Harvard team will be the first in the world to move solar geoengineering out of the lab and into the stratosphere, with a project called the Stratospheric Controlled Perturbation Experiment (SCoPEx). The first phase — a US$3-million test involving two flights of a steerable balloon 20 kilometres above the southwest United States — could launch as early as the first half of 2019. Once in place, the

Frank Keutsch, Zhen Dai and David Keith (left to right) in Keutsch's laboratory at Harvard University.

experiment would release small plumes of calcium carbonate, each of around 100 grams, roughly equivalent to the amount found in an average bottle of off-the-shelf antacid. The balloon would then turn around to observe how the particles disperse.

The test itself is extremely modest. Dai, whose doctoral work over the past four years has involved building a tabletop device to simulate and measure chemical reactions in the stratosphere in advance of the experiment, does not stress about concerns over such research. "I'm studying a chemical substance," she says. "It's not like it's a nuclear bomb."

Nevertheless, the experiment will be the first to fly under the banner of solar geoengineering. And so it is under intense scrutiny, including from some environmental groups, who say such efforts are a dangerous distraction from addressing the only permanent solution to climate change: reducing greenhouse-gas emissions. The scientific outcome of SCoPEx doesn't really matter, says Jim Thomas, co-executive director of the ETC Group, an environmental advocacy organization in Val-David, near Montreal, Canada, that opposes geoengineering: "This is as much an experiment in changing social norms and crossing a line as it is a science experiment."

Aware of this attention, the team is moving slowly and is working to set up clear oversight for the experiment, in the form of an external advisory committee to review the project. Some say that such a framework, which could pave the way for future experiments, is even more important than the results of this one test. "SCoPEx is the first out of the gate, and it is triggering an important conversation about what independent guidance, advice and oversight should look like," says Peter Frumhoff, chief climate scientist at the Union of Concerned Scientists in Cambridge, Massachusetts, and a member of an independent panel that has been charged with selecting the head of the advisory committee. "Getting it done right is far more important than getting it done quickly."

JOINING FORCES

In many ways, the stratosphere is an ideal place to try to make the atmosphere more reflective. Small particles injected there can spread around the globe and stay aloft for two years or more. If placed strategically and regularly in both hemispheres, they could create a relatively uniform blanket that would shield the entire planet (see 'Global intervention'). The process does not have to be wildly expensive; in a report last month, the Intergovernmental Panel on Climate Change suggested that a fleet of high-flying aircraft could deposit enough sulfur to offset roughly 1.5 °C of warming for around $1 billion to $10 billion per year[1].

Most of the solar geoengineering research so far has focused on sulfur dioxide, the same substance released by Mount Pinatubo. But sulfur might not be the best candidate. In addition to cooling the planet, the aerosols generated in that eruption sped up the rate at which chlorofluorocarbons deplete the ozone layer, which shields the planet from the Sun's harmful ultraviolet radiation. Sulfate aerosols are also warmed by the Sun, enough to potentially affect the movement of moisture and even alter the jet stream. "There are all of these downstream effects that we don't fully understand," says Frank Keutsch, an atmospheric chemist at Harvard and SCoPEx's principal investigator.

The SCoPEx team's initial stratospheric experiments will focus on calcium carbonate, which is expected to absorb less heat than sulfates and to have less impact on ozone. But textbook answers — and even Dai's tabletop device — can't capture the full picture. "We actually don't know what it would do, because it doesn't exist in the stratosphere," Keutsch says. "That sets up a red flag."

SCoPEx aims to gather real-world data to sort this out. The experiment began as a partnership between atmospheric chemist James Anderson of Harvard and experimental physicist David Keith, who moved to the university in 2011. Keith has been investigating a variety of geoengineering options off and on for more than 25 years. In 2009, while at the University of Calgary in Canada, he founded the company Carbon Engineering, in

Squamish, which is working to commercialize technology to remove carbon dioxide from the atmosphere. After joining Harvard, Keith used research funding he had received from Microsoft co-founder Bill Gates to begin planning the experiment.

Keutsch, who got involved later, is not a climate scientist and is at best a reluctant geoengineer. But he worries about where humanity is heading, and what that means for his children's future. When he saw Keith talk about the SCoPEx idea at a conference after starting at Harvard in 2015, he says his initial reaction was that the idea was "totally insane". Then he decided it was time to engage. "I asked myself, an atmospheric chemist, what can I do?" He joined forces with Keith and Anderson, and has since taken the lead on the experimental work.

AN EYE ON THE SKY

Already, SCoPEx has moved farther along than earlier solar geoengineering efforts. The UK Stratospheric Particle Injection for Climate Engineering experiment, which sought to spray water 1 kilometre into the atmosphere, was cancelled in 2012 in part because scientists had applied for patents on an apparatus that could ultimately affect every human on the planet. (Keith says there will be no patents on any technologies involved in the SCoPEx project.) And US researchers with the Marine Cloud Brightening Project, which aims to spray saltwater droplets into the lower atmosphere to increase the reflectivity of ocean clouds, have been trying to raise money for the project for nearly a decade.

Although SCoPEx could be the first solar geoengineering experiment to fly, Keith says other projects that have not branded themselves as such have already provided useful data. In 2011, for example, the Eastern Pacific Emitted Aerosol Cloud Experiment pumped smoke into the lower atmosphere to mimic pollution from ships, which can cause clouds to brighten by capturing more water vapour. The test was used to study the effect on marine clouds, but the results had a direct bearing on geoengineering science: the brighter clouds produced a cooling effect 50 times greater than the warming effect of the carbon emissions from the researchers' ship[2].

Keith says that the Harvard team has yet to encounter public protests or any direct opposition — aside from the occasional conspiracy theorist. The challenge facing researchers, he says, stems more from a fear among science-funding agencies that investing in geoengineering will lead to protests by environmentalists.

> "There are all of these downstream effects that we don't fully understand."

To help advance the field, Keith set a goal in 2016 of raising $20 million to support a formal research programme that would cover not just the experimental work, but also research into modelling, governance and ethics. He has raised around $12 million so far, mostly from philanthropic sources such as Gates; the pot provides funding to dozens of people, largely on a part-time basis.

Keith and Keutsch also want an external advisory committee to review SCoPEx before it flies. The committee, which is still to be selected, will report to the dean of engineering and the vice-provost for research at Harvard. "We see this as part of a process to build broader support for research on this topic," Keith says.

Keutsch is looking forward to having the guidance of an external group, and hopes that it can provide clarity on how tests such as his should proceed. "This is a much more politically challenging experiment than I had anticipated," he says. "I was a little naive."

SCoPEx faces technical challenges, too. It must spray particles of the right size: the team calculates that those with a diameter of about 0.5 micrometres should disperse and reflect sunlight well. The balloon must also be able to reverse its course in the thin air so that it can pass through its own wake. Assuming the team is able to find the calcium carbonate plume — and there is no guarantee that they can — SCoPEx needs instruments that can analyse the particles and, it is hoped, carry samples back to Earth.

"It's going to be a hard experiment, and it may not work," says David

Fahey, an atmospheric scientist at the National Oceanic and Atmospheric Administration in Boulder, Colorado. In the hope that it will, Fahey's team has provided SCoPEx with a lightweight instrument that can reliably measure the size and number of particles that are released. The balloon will also be equipped with a laser device that can monitor the plume from afar. Other equipment that could collect information on the level of moisture and ozone in the stratosphere could fly on the balloon as well.

UP TO THE STRATOSPHERE

Keutsch and Keith are still working out some of the technical details. Plans with one balloon company fell through, so they are now working with a second. And an independent team of engineers in California is working on options for the sprayer. To simplify things, the SCoPEx group plans to fly the balloon during the spring or autumn, when stratospheric winds shift direction and — for a brief period — calm down, which will make it easier to track the plume.

For all of these reasons, Keutsch characterizes the first flight as an engineering test, mainly intended to demonstrate that everything works as it should. The team is ready to spray calcium carbonate particles, but could instead use salt water to test the sprayer if the advisory committee objects.

Keith still thinks that sulfate aerosols might ultimately be the best choice for solar geoengineering, if only because there has been more research about their impact. He says that the possibility of sulfates enhancing ozone depletion should become less of a concern in the future, as efforts to restore the ozone layer through pollutant reductions continue. Nevertheless, his main hope is to establish an experimental programme in which scientists can explore different aspects of solar geoengineering.

There are a lot of outstanding questions. Some researchers have suggested that solar geoengineering could alter precipitation patterns and even lead to more droughts in some regions. Others warn that one of the possible benefits of solar geoengineering — maintaining crop yields by protecting them from heat stress — might not come to pass. In a study published in August, researchers found that yields of maize (corn), soya, rice and wheat[3] fell after two volcanic eruptions, Mount Pinatubo in 1991 and El Chichón in Mexico in 1982, dimmed the skies. Such reductions could be enough to cancel out any potential gains in the future.

Keith says the science so far suggests that the benefits could well outweigh the potential negative consequences, particularly compared with a world in which warming goes unchecked. The commonly cited drawback is that shielding the Sun doesn't affect emissions, so greenhouse-gas levels would continue to rise and the ocean would grow even more acidic. But he suggests that solar geoengineering could reduce the amount of carbon that would otherwise end up in the atmosphere, including by minimizing the loss of permafrost, promoting forest growth and reducing the need to cool buildings. In an as-yet-unpublished analysis of precipitation and temperature extremes using a high-resolution climate model, Keith and

Global intervention

One way to cool the planet quickly would be to make the sky block more sunlight. But predicting the knock-on effects — both positive and negative — remains a major challenge.

High-flying planes could release small particles into the stratosphere to reflect incoming rays. One estimate says this could reduce global temperatures by roughly 1.5 °C for less than US$10 billion a year.

Stratosphere ~10–50 km

Researchers aim to release calcium carbonate into the stratosphere next year to test some aspects of solar geoengineering.

Sunlight

Lower atmosphere

Cooler temperatures and more scattered light could promote the growth of forests and other ecosystems, locking more atmospheric carbon away.

Dimmer skies could shift global precipitation patterns, pulling water resources from some areas.

Without action on greenhouse gases, the oceans would continue to absorb carbon dioxide and grow more acidic.

Crops would benefit from reduced heat stress, but lower levels of direct sunlight could hamper growth.

others found that nearly all regions of the world would benefit from a moderate solar geoengineering programme. "Despite all of the concerns, we can't find any areas that would be definitely worse off," he says. "If solar geoengineering is as good as what is shown in these models, it would be crazy not to take it seriously."

There is still widespread uncertainty about the state of the science and the assumptions in the models — including the idea that humanity could come together to establish, maintain and then eventually dismantle a well-designed geoengineering programme while tackling the underlying problem of emissions. Still, prominent organizations, including the UK Royal Society and the US National Academies of Sciences, Engineering, and Medicine, have called for more research. In October, the academies launched a project that will attempt to provide a blueprint for such a programme.

Some organizations are already trying to promote discussions among policymakers and government officials at the international level. The Solar Radiation Management Governance Initiative is holding workshops across the global south, for instance. And Janos Pasztor, who handled climate issues under former UN secretary-general Ban Ki-moon, has been talking to high-level government officials around the world in his role as head of the Carnegie Climate Geoengineering Governance Initiative, a non-profit organization based in New York. "Governments need to engage in this discussion and to understand these issues," Pasztor says. "They need to understand the risks — not just the risks of doing it, but also the risks of not understanding and not knowing."

One concern is that governments might one day panic over the consequences of global warming and rush forward with a haphazard solar-geoengineering programme, a distinct possibility given that the costs are cheap enough that many countries, and perhaps even a few individuals, could probably afford to go it alone. These and other questions arose earlier this month in Quito, Ecuador, at the annual summit of the Montreal Protocol, which governs chemicals that damage the stratospheric ozone layer. Several countries called for a scientific assessment of the potential effects that solar geoengineering could have on the ozone layer, and on the stratosphere more broadly.

If the world gets serious about geoengineering, Fahey says that there are plenty of sophisticated experiments that researchers could do using satellites and high-flying aircraft. But for now, he says, SCoPEx will be valuable — if only because it pushes the conversation forward. "Not talking about geoengineering is the greatest mistake we can make right now." ■

Jeff Tollefson *is a reporter for* Nature *in New York City.*

1. Intergovernmental Panel on Climate Change. *Global Warming of 1.5 °C* (IPCC, 2018).
2. Russell, L. M. *et al. Bull. Am. Meteorol. Soc.* **94**, 709–729 (2013).
3. Proctor, J., Hsiang, S., Burney, J., Burke, M. & Schlenker, W. *Nature* **560**, 480–483 (2018).
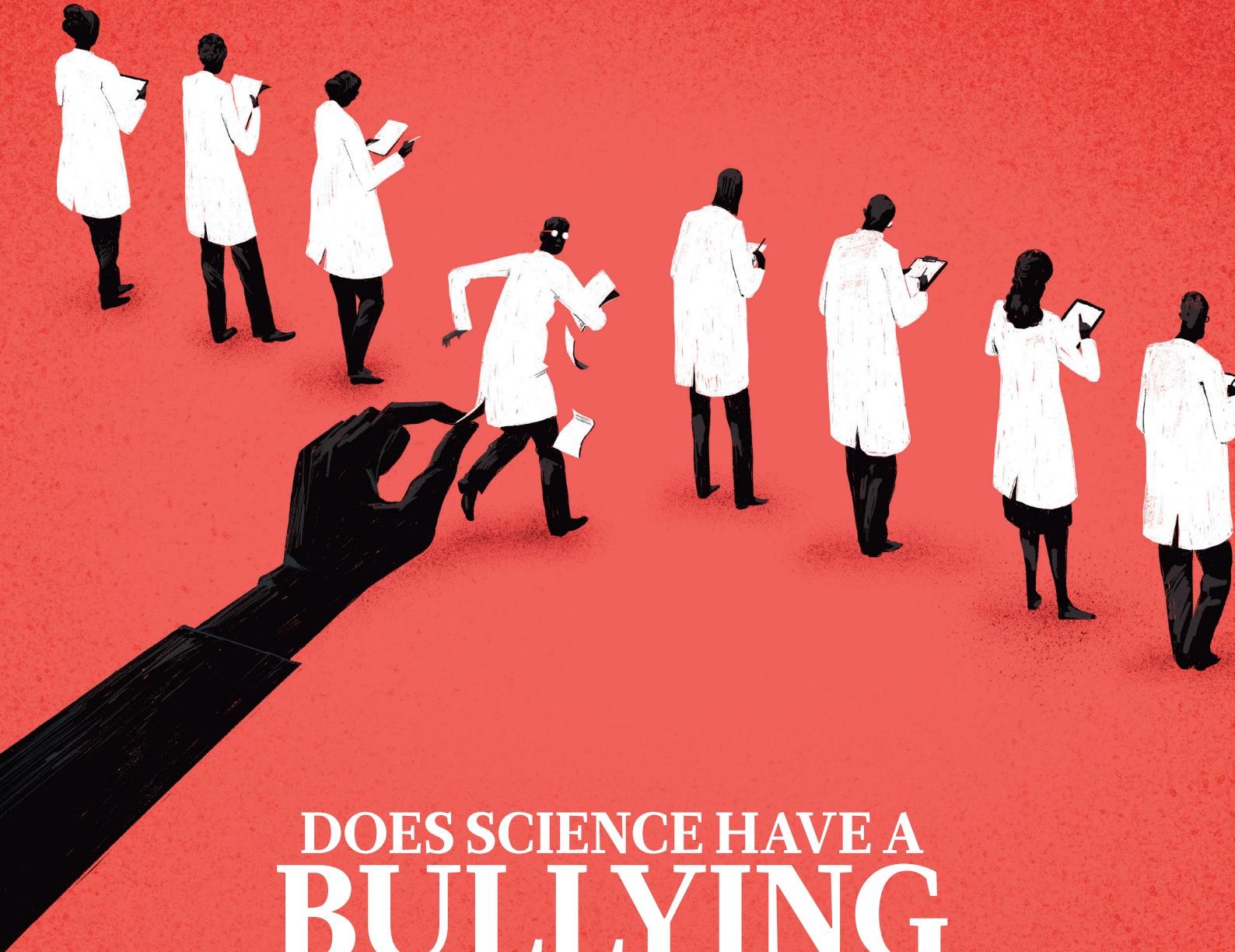
DOES SCIENCE HAVE A
BULLYING
PROBLEM?

A spate of bullying allegations have rocked some high-profile science institutions. Here's how researchers, universities and funders are dealing with the issue.

BY HOLLY ELSE

In August, accusations of bullying roiled the Institute of Cancer Research in London, one of the leading science centres in the United Kingdom. A prominent cancer researcher there, geneticist Nazneen Rahman, resigned from the institute following an investigation into allegations that she had bullied her staff. And in an unprecedented move, the biomedical charity the Wellcome Trust revoked £3.5 million (US$4.5 million) of the funding it had given her.

Three months on, many more people from Rahman's lab have left the institute. Yet most of the details about the case remain hidden from the public: Rahman has not commented about the allegations and the institute has released little information. It even withheld certain findings from the Wellcome Trust because they contained highly confidential personal information. The secrecy — and the resulting confusion — are prime examples of the difficulties that scientific institutions and researchers face in dealing with the thorny issue of bullying.

The case is part of a spate of allegations that have rocked major scientific institutions in the past year. At Germany's prestigious Max Planck Society, two directors were accused of bullying; and the UK-based Leverhulme Trust revoked £1 million in funding from palaeontologist Nicholas Longrich at the University of Bath following an investigation into bullying allegations. One of the world's leading genomics centres, the Wellcome Sanger Institute in Hinxton, UK, has also investigated claims of bullying. But the decision to clear the Sanger's management of this and other allegations has led some of those who complained to question the scope and extent of the probe. The Wellcome and other science funders, including Cancer Research UK (CRUK), have announced policies this year that prohibit bullying as well as other forms of harassment.

The flurry of activity surrounding bullying has raised questions about how scientific organizations are run and how some researchers conduct themselves. Here, *Nature* examines what constitutes bullying, why so many accusations are arising and what impact it is having on research and on those who do it.

What is bullying?

Bullying between colleagues is commonly defined by psychologists, unions and workplace scholars as repeated and malicious mistreatment of someone that results in harm. At its most obvious, this behaviour involves shouting, insulting or intimidating victims. But bullying can include more subtle actions, says Alison Antes, a workplace psychologist who studies researcher leadership and management practices at Washington University in St Louis, Missouri.

It can take the form of someone spreading malicious rumours about another, undermining their work and opinions, or withholding information necessary for them to do their jobs. Supervisors can become bullies if they are overbearing, constantly changing a person's duties or giving them impossible workloads or unachievable deadlines. These more subtle forms of malicious conduct can often cause the most problems because they tend to be difficult to detect and are open to differing interpretations, says Antes.

Some actions might fit into a grey zone. What one person considers firm management, another might consider bullying, says Antes. It is not difficult to imagine, for example, a PhD supervisor giving a student a raft of unfamiliar experiments to complete, with a deadline that leaves the student stressed and working all night. Is this bullying?

The answer depends on the broader behaviour and approach, explains Loraleigh

"THIS ISN'T PECULIAR TO SCIENCE, WE WILL SEE IT IN ALL WALKS OF LIFE."

Keashly, a communications scientist at Wayne State University in Detroit, Michigan. 'Tough' supervisors are not bullies if they set up clear expectations and communicate them directly to reportees. They will also acknowledge and appreciate staff members who meet those expectations. If employees do not achieve their goals, good supervisors will give specific and constructive feedback, she says.

Naomi Ellemers, a social psychologist at Utrecht University in the Netherlands who has studied how people are treated in academia, adds that supervisors on the right side of the line will give people the time, support and resources to achieve their goals, and treat them respectfully.

A bully, by contrast, is typically not interested in developing relationships that allow their subordinates to grow professionally, says Keashly. They might also dish out bullying behaviour on a whim, whether or not the person they are targeting has failed to perform well, she adds.

How common is it in research — and is it getting worse?

Nobody knows how much bullying goes on in science, because few people have investigated the issue. Studies of bullying in workplaces began only in the 1990s, and some researchers have yet to examine what goes on in their own back yards.

But Keashley thinks that this needs to change so that the behaviour can be better managed.

Her research, which draws on published evidence of bullying in academia from around the world, suggests that, in general, one-quarter to one-third of academics say that they have been bullied in the past year[1].

Around 40% say that they have witnessed or heard about bullying happening to someone else. This is considerably higher than the reports of bullying in the general workplace. Studies in the United States report that 10–14% of people in the general working population say that they have experienced bullying over the previous year[1].

One of the largest studies of bullying in universities — surveying 14,000 higher-education staff — was published by the UK University and College Union in 2012 (ref. 2). It found that the rate of bullying varied hugely among the 92 institutions surveyed. Between 2% and 19% of staff at each university said they were always or often subject to bullying at work.

Universities came out better than average in an earlier survey, this one published in 2000 and sponsored by the British Occupational Health Research Foundation[3], which included 5,288 workers in 15 fields. Just 7% of the 483 respondents who work in higher education say that they are occasionally or regularly bullied — the third-lowest score of all the professions looked at. (Only retailing and manufacturing has less bullying.) But Cary Cooper, a workplace psychologist at the University of Manchester, UK, who co-authored the study, says that this under-represents the true problem in universities. His survey had a relatively strict definition of bullying: workers qualified as being bullied if they had experienced persistent demeaning and devaluing treatment.

For comparison, a study of bullying in neonatal intensive-care units at 17 Greek hospitals found that more than half of the almost 400 doctors and nurses surveyed had experienced bullying[4].

And an online survey of more than 1,000 US adults conducted last April reported that 19% had experienced bullying at work[5].

Because there are so few data about bullying in research, and specifically science, Keashley and other researchers say it is not clear whether the problem is getting worse. Matt Waddup, head of policy at the UK University and College Union in London, says that bullying is not always easy to pin down in cases that come to the union, because it is often a component of other problems that members have. But he thinks it is on the rise.

Part of the reason could be that people across society are reconsidering what types of behaviours are acceptable. Ellemers says that the #metoo movement has made it a little easier for those in low-power positions to report bullying, harassment and other inappropriate behaviours exerted by those above them. It has also spurred those in charge to take action instead of dismissing or ignoring the

complaints, which often happened previously, she says.

Karen Vousden, chief scientist at CRUK in London, which recently introduced an anti-bullying policy for the labs it funds, says that society at large is now discussing these issues. "This isn't peculiar to science, we will see it in all walks of life," she says.

What contributes to bullying in science?

For the most part, says Antes, principal investigators generally "love what they do and do the right thing". But there are clearly exceptions — and certain factors in scientific research seem to encourage what some academics call abusive supervision.

Lab heads wield a lot of power over their trainees — students and postdocs — who depend on them for help, recommendations and opportunities, says Ellemers. This type of dependence and hierarchical structure can allow people to get away with bullying because it makes it difficult for those targeted or watching to confront the perpetrator, raise it with more senior colleagues or simply walk out. As a result, bullying can continue unchallenged for a long time, she says.

And bullying is not always malicious: the intense pressure to get grants, results and publications can push people to behave in problematic ways unintentionally, adds Antes.

According to another idea, science is susceptible to bullying partly because of the types of people who tend to choose that career. "In academia you do deal with a lot of individuals who are very intelligent but also have large egos," says Matthew Martin, who

What are scientific institutions doing about it?

The majority of UK universities have policies that prohibit bullying and harassment, says Waddup. These documents typically include definitions and examples and they advise on what to do if someone encounters such problems.

In July, the University of Bath reprimanded Longrich after it found that he had violated its dignity and respect policy. The institution issued him with a verbal warning and made changes to his 'supervisory arrangements'. Some people who have worked with Longrich feel that the university's initial actions did not go far enough. Subsequently, the Leverhulme Trust, which had funded Longrich, revoked his £1-million grant. Longrich has not responded to *Nature*'s repeated requests for comment. A University of Bath spokesperson told *Nature*: "Our HR procedures ensure people involved are treated reasonably, consistently and fairly."

In general, having policies is not enough, says C. K. Gunsalus, a specialist in research integrity at the University of Illinois at Urbana–Champaign. To stamp out bad behaviour, leaders need to apply policies consistently and show that bullying has consequences, she says. "One of the worst things you can do is start the process and abandon it. It reinforces the problem."

Bullying policies vary widely around the world. They are less common at universities and other institutions in the United States than in the United Kingdom. Unpublished research by Leah Hollis at Morgan State University in Baltimore, Maryland, who studies bullying in higher education, suggests that only around

most people do," she says.

Those who are bullied are more likely to be distracted and make mistakes, says Keashly. At worst, bullying can contribute to long-term problems with mental or physical health. That has an impact beyond the victim themselves, eroding the creativity, productivity and well-being of an entire lab.

After Rahman resigned and the Wellcome revoked her funding, the upheaval had ripple effects. The Institute of Cancer Research says that it followed standard processes for when a team leader leaves. Only one-third of the 15 people in her research team still work at the institute.

One concern about bullying is that it can drive people away from science permanently, especially those who were the targets, says Vousden. "Our workforce is incredibly precious. We spend huge amounts of time on mentoring and funding people," she says. "Our scientists are in some degree our most valuable component."

What needs to be done?

The next big job for institutions, says Vousden, is to create an open and supportive atmosphere in which people feel comfortable enough to bring up any concerns in a non-confrontational way. This can help to prevent situations from escalating "to the point where you have 50 people making complaints about 10 years of behaviour", she says.

Towards this goal, CRUK will be auditing the institutions it funds to check that they are adhering to its anti-bullying policy.

Another important step is for universities to offer training to scientists who assume management roles, says Cooper. Institutions should also reward researchers for taking on management tasks.

Hollis says that institutions without bullying policies should develop and put them in place. Crucially, they then need to follow the procedures. "It sounds simple, but many schools don't follow such policies," she says.

And the policies must apply "regardless of whether the bully is a vice-president or grounds workers", says Hollis. "Bullying occurs because the organization allows it to occur." ∎

"ONE OF THE WORST THINGS YOU CAN DO IS START THE PROCESS AND ABANDON IT."

studies bullying at West Virginia University, in Morgantown. And some egocentric people might be more prone to bullying because they are unconcerned with others' feelings, he proposes.

Often in science, there can be only a handful of people who are experts in a specific field, so junior researchers who experience bullying might think that it is worth putting up with the behaviour because in the long run it will pay off for them, explains Antes. "Your career success starts to be woven around their success," she says, making it even harder to speak out about poor behaviour.

And some researchers could have spent their early careers in a lab where bullying behaviour was the norm. They might be trying to use these tactics on their staff because they think that is what made them successful. So the bullying behaviours are actually coming from a place of care, says Antes — a perception that this will help others.

one-fifth of institutions have such policies that are easy to find. In France, workplace bullying is referred to as 'moral harassment' and is illegal; similar laws exist in Australia, Sweden, Belgium and several Canadian provinces.

As harassment and bullying accusations have captured more attention in the past year, several major science funders have stepped up to develop policies. The Wellcome Trust's policy specifically prohibits bullying as well as harassment. Some other funders are less clear. Neither the US National Science Foundation nor the US National Institutes of Health specifically mentions bullying in its anti-harassment policies.

What is the effect on science and scientists?

No one knows whether bullying has a negative impact on science — but Antes suspects it does. "Maybe some people can thrive in that environment, but I don't think

Holly Else *is a reporter with* Nature *in London.*

1. Keashley, L. *SPECTRA* **51**, 23–28 (2015).
2. University and College Union. *2012 Occupational Stress Survey: the Relationships stressor in HE* (UCU, 2012); available at https://go.nature.com/2s32qdt
3. Hoel, H. & Cooper, C. *Destructive conflict and bullying at work* (Manchester School of Management, 2000); available at https://go.nature.com/2zhd9kl
4. Chatziioannidis, I., Bascialla, F. G., Chatzivalsama, P., Vouzas, F. & Mitsiakos, G. *BMJ Open* **8**, e018766 (2018).
5. Namie, G. *2017 Workplace Bullying Institute U.S. Workplace Bullying Survey* (Workplace Bullying Institute, 2017); see https://go.nature.com/2kqjlsu

COMMENT

ILLUSTRATION BY DAVID PARKINS

Statistical pitfalls of personalized medicine

Misleading terminology and arbitrary divisions stymie drug trials and can give false hope about the potential of tailoring drugs to individuals, warns **Stephen Senn**.

Personalized medicine aims to match individuals with the therapy that is best suited to them and their condition. Advocates proclaim the potential of this approach to improve treatment outcomes by pointing to statistics about how most drugs — for conditions ranging from arthritis to heartburn — do not work for most people[1]. That might or might not be true, but the statistics are being misinterpreted. There is no reason to think that a drug that shows itself to be marginally effective in a general population is simply in want of an appropriate subpopulation in which it will perform spectacularly.

The reasoning follows a familiar, flawed pattern. If more people receiving a drug improve compared with those who are given a placebo, then the subset of individuals who improved is believed to be somehow special. The problem is that the distinction between these 'responders' and 'non-responders' can be arbitrary and illusory.

Much effort then goes into the effort to uncover a trait to explain this differential response, without assessing whether or not such a differential exists. I think that this is one of many reasons why a large proportion of biomarkers thought to distinguish patient subgroups fall flat.

Researchers need to be much more careful.

To be clear, I am not talking about research, often in cancer, that defines subpopulations of patients in advance. In that scenario, the aim is to test prospectively whether a particular drug works better (or worse) in people whose cancer cells have a specific genetic defect — a biomarker such as a *HER2* mutation in breast cancer or the *BCR–ABL* fusion gene in leukaemia. (It's worth stating that the overall percentage of US patients with advanced or metastatic cancer who benefit from such 'genome-informed' cancer drugs is estimated to be less than 7% at best[2]; the proportion is likely to be lower for those ▶

whose cancer is at an earlier stage.)

What I take issue with is the de facto assumption — often made in studies of chronic diseases such as migraine and asthma — that the differential response to a drug is consistent for each individual, predictable and based on some stable property, such as a yet-to-be-discovered genetic variant.

Consider an actual clinical trial in which 71 patients were treated with two doses. Twenty 'responded' to both doses, 29 to neither dose and 14 to the higher dose, but not the lower one. That is as expected. More surprising is that eight 'responded' to the lower dose and not the higher one, which is at odds with how drugs are known to work. The most likely explanation is that the 'response' is not a permanent characteristic of a person receiving the treatment; rather, it varies from occasion to occasion. In this example, the fact that two doses of the same drug were being compared alerts us to the need to consider that source of variability. If the comparison instead involved different molecules, researchers might then overlook the explanation of occasion-to-occasion variation and jump to the conclusion that the results must reflect a differential response.

I have seen unsubstantiated interpretations waft through the literature. They start with trials designed to show whether a drug works, and then get misinterpreted. For example, a 2005 study found that one ulcer treatment led to healing in 96% of patients after 8 weeks, and another treatment healed 92% of patients, a difference of 4% (ref. 3). This finding filtered into a 2006 meta-analysis[4], and then a third article[1] followed an all-too-common statistical practice, stating that only 1 in 25 (or 4%) of patients would benefit from the first ulcer treatment. It is not hard to imagine other researchers carrying out futile work to try to understand why.

TRIAL TRAPS
Here are some common pitfalls.

Lazy language. Participants in clinical trials are often categorized as being responders or non-responders on the basis of an arbitrary measure of improvement — such as a certain percentage drop in established clinical scales that assess depression or schizophrenia. It does not necessarily follow that any individual who improves owes that improvement to the treatment. Researchers who acknowledge in the methods section of a paper that an observed change is not a proven effect of a drug often forget to make that distinction in the discussion. Variations are uncritically attributed to characteristics of the person receiving treatment rather than to numerous other possibilities.

Arbitrary dichotomies. Other classifications can depend on whether a participant falls on one side or another of a boundary on

COMPARE EACH PATIENT AT LEAST TWICE
To find out whether a drug works better for some people, researchers can compare it to a placebo in those individuals more than once. (All data here are simulated.)

JUST ONE TEST: DOES THE DRUG WORK?
A single comparison to a placebo that gives results such as these suggests only that, overall, the drug works better than the placebo.

TWO TESTS: DOES THE DRUG WORK BETTER FOR SOME?
Many comparisons can show whether individuals respond in the same way to a drug each time.

DID A PATIENT RESPOND TO TREATMENT?
+ Yes, to both comparisons   × Yes, but just in one   ○ No

CONSISTENT RESPONSE
There is potential for tailored therapy.

INCONSISTENT RESPONSE
Futile to try to pinpoint a subset of responders.
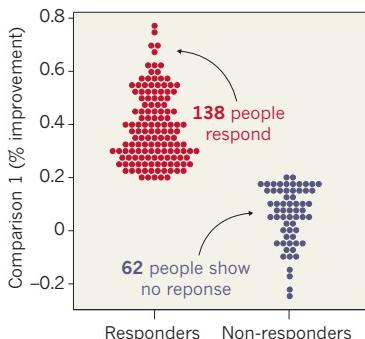
a continuous measurement. For example, a person with multiple sclerosis who relapsed 364 days after treatment is a non-responder; one who relapses 365 days after treatment is a responder. This is simplistic — it recasts differences of degree as differences of kind. Worse, it causes an unfortunate loss of information, and means that clinical trials must enrol more participants than would otherwise be needed to reach a sound conclusion[5,6].

Participants' variability. Physiology fluctuates. Trial participants are often labelled as responders after one measurement, post-treatment, with the tacit assumption that the same treatment in the same person on another occasion would yield the same observation. But repeated observations of the same person with a disease such as asthma or high blood pressure show that the result after treatment can vary.

Inappropriate yardsticks. Judging whether a drug works depends on making assumptions about what would have happened without the treatment — a counterfactual. One common technique for estimating the counterfactual is to take baseline measurements; for instance, the volume of air that people with asthma can force from their lungs in one second at the start of a trial. But baselines are a poor choice of counterfactual. Guidelines agreed by drug regulators in the European Union, Japan and the United States disparage their use as controls.

There are many reasons besides treatment — such as regression to the mean or variation in clinical settings — that might explain a difference from baseline, especially if measurements such as elevated blood pressure or reduced lung capacity are used to determine who can enrol in a clinical trial. Let's say Patient X was enrolled in a trial after meeting the criteria for having a blood-pressure
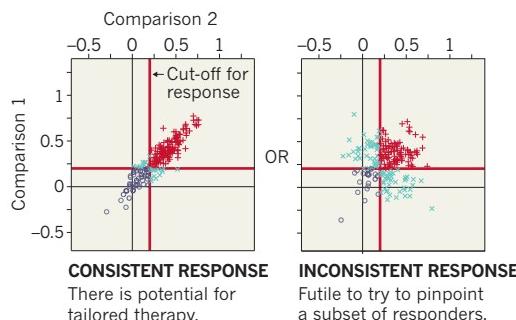
measurement of more than 130/90 mm Hg. She is given a drug, after which her blood pressure measures 120/80 mm Hg. One possibility is that the drug affected her blood pressure. Another is that 125/85 mm Hg (or some other intermediate value) is her mean blood pressure, and that she had a bad day on enrolment and a good day later. Yet another possibility is that her blood pressure was measured at different times of the day, at different places or by different people.

For measurements such as pain scores and cholesterol levels, predictions for individuals — based on an average of all participants — can be more accurate than predictions based on an individual's own data taken just once[7].

Rates of response. Suppose that in a large trial for an antidepressant, 30% of patients have a satisfactory outcome in terms of their score on the Hamilton Depression Rating Scale after taking a placebo, and 50% show a satisfactory outcome after taking the drug. This means that the probability of a good outcome observed with the drug is 20% higher than with the placebo. Or put another way, on average, if five patients were treated with the drug, one more would experience a satisfactory outcome. This statistic is an example of what is called the 'number needed to treat' (NNT).

This concept was introduced 30 years ago[8] and is extremely popular in evidence-based medicine and assessments of health technology. Unfortunately, NNTs are often falsely interpreted. Consider a trial comparing paracetamol to a placebo for treating tension headache. After 2 hours, 50% of people treated with the placebo are pain-free, as are 60% of those who were treated with paracetamol. The difference is 10% and the NNT is 10. However, if paracetamol works for 100% of participants in 60% of the times they are treated, it will give the same NNT as if it works for 60% of the participants 100% of the time.

A high NNT should not be taken to imply that a drug works really well for a specific, narrow subset of people. It could simply mean that a drug is just not that effective across all individuals.

Subsequence, not consequence. All of the errors discussed so far lead to the assumption that what has happened, for good or ill, has been caused by what was done before — that if a headache disappeared, it was because of the drug. It is ironic that the evidence-based-medicine movement, which has done so much to enthrone the randomized clinical trial as a principled and cautious way of establishing causation across populations, consistently fails to establish causation in the context of personalized medicine.

WAY FORWARD

These warnings are not intended to discourage researchers from pursuing precision medicine. Rather, they are meant to encourage them to get a better sense of its potential at the outset.

How to improve? One thing we need more of are *N*-of-1 trials. These studies repeatedly test multiple treatments in the same person, including the same treatment multiple times (see 'Compare each patient at least twice').

With such designs, we can assess differences between the same drug being administered on many occasions, and compare those data with differences seen when different drugs are administered in the same way. They are being used, for example, in trials of fentanyl for pain control in individuals with cancer[9] and of temazepam for people with sleep disturbances[10].

When medicines are given on many occasions for a chronic or recurring condition, *N*-of-1 studies are a good way of establishing the scope for personalized medicine[11]. When drugs are given once or infrequently for degenerative or fatal conditions, careful modelling of repeated measures can help. Whatever their approach, trial designers must hunt down sources of variation. To work out how much of the change observed is due to variability within individuals requires more careful design and analysis[12].

Another advance would be to drop the use of dichotomies[5]. Statistical analysis of continuous measurements is straightforward but underused. More-widespread uptake of this approach would mean that clinical trials could enrol fewer patients and still collect more information[6].

Perhaps the most straightforward adjustment would be to avoid labels such as 'responder' that encourage researchers to put trial participants in arbitrary categories. An alternative term — perhaps 'clinical improvement' or 'satisfactory endpoint' — might help. Better still, sticking with the actual measurement would reduce the peril of all the pitfalls mentioned here.

> *"Whatever their approach, trial designers must hunt down sources of variation."*

It has been a long, hard struggle in medicine to convince researchers, regulators and patients that causality is hard to study and difficult to prove. We are in danger of forgetting at the level of the individual what we have learnt at the level of the population. Realizing that the scope for personalized medicine might be smaller than we have assumed over the past 20 years will help us to concentrate our resources more carefully. Ironically, this could also help us to achieve our goals. ■

Stephen Senn *was formerly head of the Competence Center for Methodology and Statistics at the Luxembourg Institute of Health.*
e-mail: stephen@senns.demon.co.uk

1. Schork, N. J. *Nature* **520**, 609–611 (2015).
2. Marquat, J., Chen, E. Y. & Prasad, V. *JAMA Oncol.* **4**, 1093–1098 (2018).
3. Labenz, J. *et al. Aliment. Pharmacol. Ther.* **21**, 739–746 (2005).
4. Gralnek, I. M., Dulai, G. S., Fennerty, M. B. & Spiegel, B. M. *Clin. Gastroenterol. Hepatol.* **4**, 1452–1458 (2006).
5. Fedorov, V., Mannino, F. & Zhang, R. *Pharm. Stat.* **8**, 50–61 (2009).
6. Senn, S. *Stat. Biopharm. Res.* **5**, 204–210 (2013).
7. Senn, S. *Br. Med. J.* **329**, 966–968 (2004).
8. Laupacis, A., Sackett, D. L. & Roberts, R. S. *N. Engl. J. Med.* **318**, 1728–1733 (1988).
9. Portenoy, R. K., Burton, A. W., Gabrail, N., Taylor, D. & Fentanyl Pectin Nasal Spray 043 Study Group. *Pain* **151**, 617–624 (2010).
10. Wegman, A. C. *et al. Fam. Pract.* **22**, 152–159 (2005).
11. Araujo, A., Julious, S. & Senn, S. *PLoS ONE* **11**, e0167167 (2016).
12. Senn, S. *Stat. Med.* **35**, 966–977 (2016).

Why academic freedom is needed more than ever

For a century, the Haldane principle has enabled government scientists to speak truth to power without fear of retribution — cherish it, urges **Ehsan Masood**.

One hundred years ago this month, shortly after the guns of the First World War fell silent, a German-speaking Scottish lawyer-turned-politician sent an 80-page report to his prime minister. In it was an idea whose echo still shapes the way in which many nations fund research — an idea arguably as important to the soul of modern science as the secular state is to modern democracy.

That idea has come to be called the Haldane principle, after its proponent, Richard Burdon Haldane. This principle says that scientists should mostly be left alone to decide which research projects should receive government funding[1–3]. (It is not to be confused with the rule about speciation, formulated by evolutionary biologist J. B. S. Haldane.) In many nations, the Haldane principle is near-totemic — regarded as the scholar's last defence against more powerful interests.

But the definition used today does not reflect the depth of vision in the original. Haldane argued in his 1918 report[4] that politicians need to do more than stay out of funding decisions. He urged them to listen to expertise, and to take time to think and reflect before reaching a conclusion. And he wrote that politicians who ask scientists for advice should resist telling them what that advice should be.

The difference matters. Today, from Istanbul to Islamabad, from Rome to Rio de Janeiro, a parade of authoritarian leaders is advancing policies that fly in the face of evidence — on energy, emissions, the environment, economics, immigration and more. Worse, these leaders are demanding that academics march to the beat of their drums.

Even in seemingly healthy democracies, the direction of travel is unmistakable. In the United Kingdom last year, a 'Haldane principle' was passed into law for the first time — but as part of a package of measures that saw universities lose the protection ▶

of the royal charters that have enshrined scholarly autonomy for centuries[5].

Today, it is researchers who are demanding protection for their ability to speak truth to power. What is remarkable about Haldane's incarnation of this idea is that it originated, not from scientists, but from the heart of government in the darkest of times.

A NEW WORLD

Britain in 1918 was a different place. The small island nation controlled about one-fifth of the world's population and one-quarter of global land area from Canada to New Zealand, including many nations in Africa, Asia and the Middle East.

The First World War had been a wake-up call. In 1917, Haldane was asked by the UK government to chair the Orwellian-sounding Machinery of Government Committee. Its remit was to re-engineer Britain's ministries to cope with peacetime, the rising powers of Germany and the United States, and the citizens of various colonial states, who would soon be demanding freedom. The seven-member committee included Edwin Montagu, secretary of state for India, and the economist and social reformer Beatrice Webb, a co-founder of the London School of Economics and Political Science.

> *"A politician argued persuasively for a check on the power of the very corridors he walked."*

Haldane was an unusual politician but the ideal chairperson for this task. He had top-level experience of government as Lord Chancellor, the person responsible for the country's judiciary. And he was available: he'd lost the post of minister for war in 1915 after a campaign in the popular press painted him as a German sympathizer. He had praised the organization of education and science in Germany, noting how its culture, industry, research and policymaking came together. For this, the newspapers called him an enemy of the British state. He was doorstepped by journalists and insulted on the street[6].

Haldane could just as easily have worked in academia. He studied philosophy at the University of Göttingen in Germany and wrote several books on the topic, including *The Reign of Relativity* in 1921, about the implications of Albert Einstein's physics. But his heart lay in public policy. He was an early advocate of expanding education and, after he left the government, helped to create a wave of civic universities in cities including Birmingham, Bristol, Leeds, Liverpool and London[6].

THINK THEN DO

The Haldane report's key recommendations included something that we take for granted today: cabinet-level ministries for health and education. In another innovation that is also now mainstream, Haldane advised that these ministries would need access to the best available advice. For example, an education ministry would need counsel from experts in childhood development, and a health ministry would need guidance from scientists working on infectious diseases. His ideal ministers were people whose time was freed from operational matters to be able to think and plan.

The most radical suggestion in the report was for an entirely new ministry of "research and information". Haldane dared to suggest that its leader should be not a party politician (the convention then, as now), but "essentially a trained thinker".

The report envisaged this ministry as a blend of government think tank and research funder. It urged that "better provision should be made for enquiry, research and reflection before policy is defined and put into operation".

The historian David Edgerton has rightly pointed out that the original report does not mention a 'Haldane principle' (see go.nature.com/2qybjbn). So where did the moniker arise? In an unpublished memo written probably in February or March of 1918, six months into the inquiry, Haldane mentions three "principles" for reorganizing government[7]. The first — a "new principle to be recognised as fundamental" — is for government and policymakers to develop "a habit of mind, a disposition to insist on the systematic study of questions before [policy] action is taken". (The other two focused on the rationale for different ministerial jobs and better financial accountability from government departments.) The memo's tone is much more direct than that of the final report, suggesting that its intended audience was probably Prime Minister David Lloyd George.

In words that ring true today, Haldane adds: "A Prime Minister is chosen as the leader of the nation largely because of his gifts as its spokesman … But he has to shape policy, and to this end requires the most highly skilled assistance, if he is not to be a bungler."

Despite this progressive thinking, there is no sugar-coating the fact that Haldane was an imperialist[8]. The needs of the British Empire were a strong factor in his calculations for science in government. There were railways to be built, botanical and geological surveys to be done, new languages and legal systems to be mastered — and catastrophic famines and outbreaks to be tackled, notably in India. All of this demanded engineers and scientists[9].

Haldane's wish for an over-arching ministerial research department never materialized. It is a brave government that would prioritize study, thought and reflection in the making of policy. But traces of the Haldane ideal can be seen in what was to follow.

His ideas are reflected today in the work of the scientists attached to the ministries dedicated to science, technology, innovation and higher education. These are largely responsible for organizing and funding teaching and research in universities and in public laboratories. They also seek expert counsel. In a few countries — notably Germany and the United Kingdom — they are also involved in industrial policy.

The ideal of independence also informs the work of chief science advisers, whose offices might be attached to those of heads of government or to departments from food to forestry, transport to trade. Since 2014, they

Students in Budapest protest in April 2017 over government interference in universities.

Richard Burdon Haldane advised in 1918 that governments need access to the best expert advice.

have been part of the International Network for Government Science Advice (INGSA), created to hone practice. The difference is that Haldane wished for such expertise to operate closer to the apex of government, and to be accountable to Parliament.

INDEPENDENCE DAY

When Haldane's report landed on the prime minister's desk, it had little impact: the end of the First World War was a busy time for statecraft. There were peace treaties to be agreed and a domestic economy to be steadied. The Ottoman empire was collapsing, and Britain and France were competing for influence in its former territories.

It was in the years during and after the Second World War that Haldane's idea of independent advice resurfaced. Scientists and engineers from many countries had created the technologies that were crucial to the Allies' victory, such as radar and the atomic bomb. These needed a degree of operational distance from politicians — a hard-won achievement, as writer C. P. Snow describes entertainingly in his 1961 book *Science and Government* (Harvard Univ. Press; see also J. Baker *Nature* **459**, 36–39; 2009).

US scientists who had held prominent policy roles during the Second World War — such as Vannevar Bush — spied an opportunity. Bush's post-war report, *Science: The Endless Frontier*, was an appeal to US leaders that if scientists could help to win the war, they could also help to hold the peace[10]. Bush

noted that they would need federal funding and, crucially, would require politicians to stay at arms' length.

And so it proved. Year after year, when governments respected the independence of the scientists they tapped for advice, the results were genuinely world-changing. Examples include the first generation of scientists who created Green Revolution agricultural technologies in the 1950s and 1960s, and the researchers whose findings led to the Montreal Protocol to protect the ozone layer in 1987. The Kyoto climate protocol of 1997 was a direct result of the efforts of the Intergovernmental Panel on Climate Change, whose members, although nominated by governments, fight hard to work without their paymasters peeping through the keyhole[11].

But Haldane's world of the honest broker starts to break down when governments stop keeping their side of the bargain.

BEWARE BUNGLERS

That is what is happening now, as an expanding network of populist political movements derides independent scholarship. For instance, Britain's staunchest supporters of the campaign to leave the European Union ('Brexiteers') disdained expert warnings of the economic and environmental costs. In his election campaign, Brazil's new president, Jair Bolsonaro, pledged to roll back the country's historical commitments on deforestation and climate change. And last month, Michael Ignatieff, rector of the Central European

University in Budapest, announced that the university will be relocating to Vienna because of sustained interference in its operations by Hungary's right-wing government.

Meanwhile, some scientists are so concerned by the ransacking of the US Environmental Protection Agency (EPA) by President Donald Trump's White House that they have reportedly set up a shadow EPA in preparation for the next administration, so that valuable knowledge isn't lost. And in Australia, former education minister Simon Birmingham was unapologetic when it emerged that he had vetoed 11 grants worth Aus$4.2 million (US$3 million) that had been cleared for funding by the Australian Research Council.

There are other examples, and there will be more as populism strengthens its grip on those who suffered as a result of the 2008 financial crisis. And that is what makes the original Haldane report a remarkable document, worth recalling now. With national security under threat, Haldane's committee could have demanded fealty from scientists and engineers. It could have insisted on ideological litmus tests. It did no such thing.

Today, more than ever, the authentic Haldane principle — and its origin story — must be cherished. In a world laid waste by war, a politician argued persuasively for a check on the power of the very corridors he walked. Haldane died in 1928, having no inkling that his Machinery of Government report would be talked of a century later. Its lasting legacy is the insight that the truth, often expendable in politics, must not be so in science advice. ∎

Ehsan Masood *was editor of* Research Fortnight *from 2009 to 2017. His latest book is* The Great Invention: The Story of GDP and the Making and Unmaking of the Modern World *(Pegasus, 2016).*
e-mail: ehsan.h.masood@gmail.com

1. UK Government. *Higher Education and Research Act 2017* Ch. 29 (Part 3: Research) (HM Stationery Office, 2017); available at https://go.nature.com/2dnyxo2
2. Willetts, D. *A University Education* (Oxford Univ. Press, 2017).
3. Innovation, Universities, Science and Skills Committee. *Eighth Report: Putting Science and Engineering at the Heart of Government Policy* Ch. 4 (House of Commons, 2009); available at https://go.nature.com/2qbzdhs
4. Machinery of Government Committee. *Ministry of Reconstruction: Report of the Machinery of Government Committee* (HM Stationery Office, 1918); available at https://go.nature.com/2q480sm
5. *Nature* **538**, 5 (2016).
6. Vincent, A. *J. Hist. Ideas* **68**, 157–179 (2007).
7. Haldane, R. B. 'Memorandum by Lord Haldane' Parliamentary Archives: W/3 – (32) P/74/12/1 (1918).
8. Haldane, R. B. *Richard Burdon Haldane: An Autobiography* (Hodder and Stoughton, 1929).
9. Baber, Z. *The Science of Empire: Scientific Knowledge, Civilization and Colonial Rule in India* (State University of New York Press, 1996).
10. Bush, V. *Science: The Endless Frontier* (US Government Printing Office, 1945).
11. Leggett, J. *The Carbon War: Global Warming and the End of the Oil Era* (Routledge, 2001).

**Which book will be a bestseller? Sales are often driven by word of mouth rather than quality.**

The secrets of status

There are laws underlying success and fame. **Mark Buchanan** savours a study of them.

A few years ago, physicist César Hidalgo and his team devised a way to rank the most famous people of all time. The criterion used for the Pantheon project was the number of languages a person's Wikipedia pages appear in. The most famous musician? Jimi Hendrix. The most famous American? Martin Luther King Jr. Perhaps inevitably, the classical Greek philosopher Aristotle heads the entire list.

Among celebrities, reality-television star Kim Kardashian comes 14th, although her fame clearly outweighs any recorded achievement. Such paradoxical mismatches of fame and attainments, as physicist and network scientist Albert-László Barabási writes in *The Formula*, reflect deep social laws that can be understood through science. Success and recognition in many realms have only a tenuous link to effort, skill or inherent excellence. Often, they are determined by less obvious factors of human behaviour that influence how attention flows through social networks.

The Formula is a fun, fast, first-hand account of efforts to use big data to pull back the curtain on our collective dynamics. As Barabási shows, hidden statistical and multiplicative processes have huge consequences in the human world, yet often operate outside our general awareness.
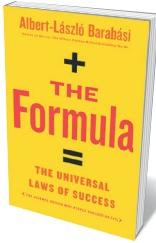
The Formula: The Universal Laws of Success
ALBERT-LÁSZLÓ BARABÁSI
Little Brown (2018)

Many of the effects described wouldn't be surprising in a system of inanimate matter, such as cracks growing irregularly in a brittle material. It would just be ordinary statistical physics. But these unstable cascades are eye-opening when documented in systems of free-willed people, where they have major impacts on people's lives. They might propel one individual vastly beyond another of identical skill, or drive rapid, unpredictable transformations of social norms such as smoking.

Barabási describes how exploiting big data collected from the web, including social media and other digital repositories, is helping researchers to tease out how success and performance actually relate to one another. If one song is more popular, or one person more wealthy, is it because of inherent differences, or merely the result of luck and random amplification? Even 20 years ago, such questions were subject only to ideological debate, not scientific

exploration. Science has changed the game.

Barabási considers puzzles in music, science, wealth, sports and wine. In some arenas, such as competitive tennis, skill and prowess are decisive. In others — including whether a book is pulped or becomes a bestseller — quality seems to be overwhelmed by network effects, such as the tendency to flock towards books that have already sold well. An intriguing example is German pilot Manfred von Richthofen (the 'Red Baron'). Remembered as one of the First World War's top flying aces, he was objectively outperformed on many counts by an almost-forgotten French counterpart, René Fonck. Von Richthofen's lasting celebrity arose, Barabási shows, in part from his early death as a war hero, which made him useful for propaganda.

How it works in general, Barabási suggests, is now becoming clear owing to the emergence from research of a number of simple "laws of success".

The first is that "performance drives success, but when performance can't be measured, networks drive success". With competitive tennis, better athletes win repeatedly, showing superiority. But when judging wine, it's not easy to find objective means of ranking: repeated blind tastings, even among wine experts, lead to wildly fluctuating outcomes. When quality is hard to measure,

observed differences in success — judged by popularity or sales, for example — follow from network effects. People rush to buy an early leader, swayed by the mistaken belief that others' choices tell them about standard.

This results in huge differences in outcome that have nothing at all to do with quality. That phenomenon is the subject of the second law: "Performance is bounded, but success is unbounded." Take the top 100 wines entered into a competition. Their true differences in quality, for example in clarity or varietal character, are generally small: they're all produced by top winemakers using similar technology. Yet one wine, because of the amplifying power of social networks, might enjoy orders of magnitude more sales than others.

Social scientists have known about such effects for decades, although research by many, including Barabási and his students, has brought them into much clearer focus. *The Formula* also covers nuanced studies showing how success can be predicted. So the third law that Barabási describes is: "Previous success × fitness = future success." Careful studies — for example, by network scientist Manuel Cebrian or complexity scholar Dashun Wang — have found that it's possible to identify how much of a product's popularity depends on its quality or fitness, and how much can be traced to random amplification by network effects. Detailed data on consumer ratings and sales over time gathered by a site such as Amazon.com can be used to disentangle herding effects (those tending to push the already popular towards further popularity) from real consumer preferences based on true perceived quality. This understanding can be used to forecast trends, but also to boost sales or consumer satisfaction.

And then there's the fourth law: "While team success requires diversity and balance, a single individual will receive credit for the group's achievements." Analyses of highly successful teams in science or business show that which individuals get the most credit has nothing to do with who actually did the work. Credit is based on perception, and is a collective social phenomenon. Effective teams absolutely require diversity, but society singles out lone individuals for the accolades.

Altogether, *The Formula* offers a rich tour of research on how relatively simple feedback forces channel our lives in surprising and counter-intuitive ways. We might think that success ought to be determined by a person's skill and hard work. Yet, more important is how other people respond, by interacting in complex social networks. Even individual success is a thoroughly social matter. ∎

Mark Buchanan *has written many books about network effects. He is based in Abergavenny, UK.*
e-mail: buchanan.mark@gmail.com

Books in brief
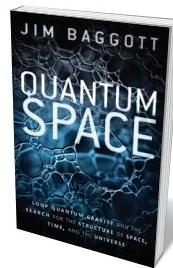
Quantum Space
Jim Baggott OXFORD UNIV. PRESS *(2018)*
Prolific physics writer Jim Baggott is back with a terrific page-turner on loop quantum gravity (LQG) — the theory posited as a solution to that chasm in physics between quantum mechanics and the general theory of relativity. Baggott digs into the how and why of what LQG might reveal about "space, time and the universe", tracing its evolution through the work of Abhay Ashtekar, Lee Smolin, Carlo Rovelli and others, to its current implications for, say, the physics of black holes. Baggott masterfully tenderizes the scientific chewiness and is careful not to over-egg what is, after all, a work in progress.
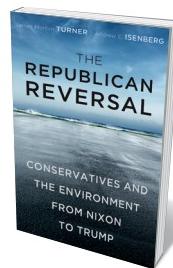
The Republican Reversal
James Morton Turner and Andrew C. Isenberg HARVARD UNIV. PRESS *(2018)*
In the 1960s and 1970s, the US Republican Party — pressured by the era's environmental movement — created the Environmental Protection Agency (EPA) and extended the Clean Air Act. Today, it busily eviscerates the EPA while denouncing climate change as a hoax. Environmental historians James Turner and Andrew Isenberg follow this reversal from Ronald Reagan's presidency on, revealing how conservative ideologues hostile to science and bent on deregulation have gradually bolted US exceptionalism to anti-environmentalism. Searingly timely and cautiously hopeful.
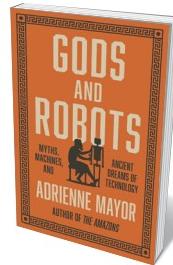
Gods and Robots
Adrienne Mayor PRINCETON UNIV. PRESS *(2018)*
More than two millennia before today's explosion in robot manufacture, bards and philosophers toyed with the concept of imitating life. Classics scholar Adrienne Mayor's astonishing chronicle harks back to mythic automata, such as "evil fembot" Pandora and bronze giant Talos. And she examines real mechanical devices — flying doves, bellowing statues and gliding Buddhas — devised by virtuosic technicians from the Mediterranean to China. A third-century BC colossus crafted for Egyptian monarch Ptolemy II Philadelphus, for instance, could stand up, sit down and pour milk.

Dark Commerce
Louise I. Shelley PRINCETON UNIV. PRESS *(2018)*
Illicit trade in human organs, wildlife, arms and rare woods has vastly expanded over the past three decades as communications and digitization have improved apace. Here, Louise Shelley, a leading researcher in the field, examines organized crime over four millennia. She unpeels its disturbing dynamics today through case studies such as Silk Road, a vastly lucrative cybersupermarket, and the much-documented illegal market in rhino horn (currently priced at US$60,000 per kilo). And she lucidly lays out the dark economy's planetary costs, as it escalates biodiversity loss and deforestation.
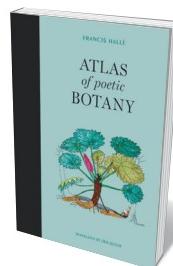
Atlas of Poetic Botany
Francis Hallé with Éliane Patriarca, transl. Erik Butler MIT PRESS *(2018)*
From the epiphytic 'hanging' plant *Guzmania lingulata* to the mushroom mimic *Helosis cayennensis,* compelling oddities crowd equatorial forests. Botanist Francis Hallé celebrates their spectacular weirdness in this sprightly homage, translated from French by Erik Butler. Alongside descriptions of clonal forests, underground trees and 'dancing' plants, Hallé sets playful stylized drawings explicating the strange behaviours, adaptations and coevolution of each species. It's a vegetal parade that reminds us, yet again, how some chunks of Earth's biosphere still smack of *terra incognita*. Barbara Kiser

Correspondence

Heat and soil vie for waste

The Renewable Heat Incentive scheme in the United Kingdom has led some industries to turn much of their organic waste into feedstocks for energy production. Previously, this material — such as biosolids from water processing — was used to replenish soil carbon. This switch might inadvertently undermine the global initiative to increase soil carbon to mitigate climate change, which was ratified in 2015.

This potential conflict needs to be part of the policy discourse and research agenda for energy and the environment. Reductions in organic carbon hamper soil functions, such as water-holding capacity, and reduce crop productivity. But knowledge gaps remain about the necessary level, composition and quality of carbon required to maintain and improve soil health and fertility.

The products of anaerobic digestion of organic wastes for heat production are widely used as soil conditioners. But they have a lower ratio of carbon to nitrogen than do conventional composted vegetation and food waste. It is not known just how low the ratio of such wastes can go before it becomes worthless to return them to the soil.

Karen L. Johnson * *Durham University, UK. *On behalf of 4 correspondents (see go.nature. com/2dxkpus for full list). karen.johnson@durham.ac.uk*

Educate public about gene–edited crops

Crop varieties created by gene editing could benefit farmers in developing countries by providing bigger yields with better nutrition and greater tolerance to stress. But the public's suspicion and fear impede the application of plant biotechnology in regions where it would be most useful. International outreach efforts are gearing up to increase public understanding of the scientific principles behind the technology. This will help governments to make informed decisions about gene-edited crops.

For example, secondary-school programmes run by universities in Malaysia and Ghana are educating the farmers, researchers and leaders of the future. Uganda's Biosciences Information Centre targets smallholders. And in the United States, Iowa State University's Plant Breeding Education in Africa Programme provides free e-learning courses to universities in sub-Saharan Africa on the application of biotechnology and genomics in plant breeding.

To increase this type of outreach, governments and development organizations need to invest in universities and secondary-school teachers, and provide them with the necessary resources.

Walter P. Suza * *Iowa State University, Ames, Iowa, USA. *On behalf of 5 correspondents (see go.nature.com/2qhy2cb for full list) wpsuza@iastate.edu*

Rat and bat hunt helps heal old rift

In the British Solomon Islands Protectorate in 1927, a warrior named Basiana led the Kwaio resistance against colonial rule of the island of Malaita, in which 15 people — including an Australian and a Briton — were killed with spears and a few rifles. The London Colonial Office asked Australia to quell the 'uprising'. In the months that followed, Australians and Solomon Islanders killed at least 60 Kwaio, desecrating shrines and violating cultural taboos. Eventually, Basiana surrendered and was hanged with six conspirators. For almost a century, these events have held back the Kwaio people, shaping their relations with 'Europeans'.

In 2015, however, the Kwainaa'isi Cultural Centre and the Australian Museum began a collaboration in East Kwaio to search for two undescribed mammals — a giant rat called *kwete* (probably *Uromys* or *Solomys*), and a monkey-faced bat (*Pteralopex*). When unresolved tensions were threatening the safety of the personnel involved, Kwaio leadership saw the relationship developed with the Australian Museum as an opportunity for reconciliation.

In July, we and other representatives and descendants of tribes and Australians affected in 1927 met in the mountains of Malaita for traditional ceremonies, exchanging pigs and shell money to resolve the dispute. The watershed event has established us as genuine partners and is a beginning to peace among Kwaio tribes, Malaita, the Solomon Islands and ultimately with Britain.

Our experience shows that all parties can benefit from biodiversity surveys if they respect local cultural processes and are built on mutual collaboration.

Esau Kekeubata * *Kwainaa'isi Cultural Centre, Kwainaa'isi, East Kwaio, Malaita, Solomon Islands.* **Tyrone Lavery** * *Biodiversity Institute, University of Kansas, Lawrence, Kansas, USA. *On behalf of 6 correspondents (see go.nature.com/2zsd2x4 for full list). tlavery@fieldmuseum.org*

Reproducibility in public and private

The reproducibility crisis in biomedical science seems to have alarmed industry more than the academic community (see C. G. Begley and L. M. Ellis *Nature* **483**, 531–533; 2012). In our view, this is because they have different yardsticks for success in research.

Despite the advent of important new therapeutics, the number of innovative treatments reaching the patient is disappointingly low. To help rectify this, industry is investing in drug-discovery alliances with peers and academic groups, and in precision medicine. It sees high standards of research quality as the route to the most promising drug candidates and to maximum return on investment.

By contrast, academic scientists may be reluctant to devote extra time and effort to confirming research results in case they fail. That would put paid to publication in high-impact journals, damage career opportunities and curtail further funding. Evidence of questionable practices such as selective publishing and cherry-picking of data indicates that rigour is not always a high priority.

Paradoxically, the impact of high standards on research objectives is different in industry and in academia. If ignored, this paradox could endanger future collaborations between scientists in the private and public sectors.

Anton Bespalov * *Heidelberg, Germany.* **Adrian G. Barnett** *Queensland University of Technology, Brisbane, Australia.* **C. Glenn Begley** * *BioCurate, Melbourne, Australia. *Competing interests declared (see go.nature.com/2retftw for details). anton.bespalov@paasp.net*

Osamu Shimomura

(1928–2018)

Chemist who illuminated bioluminescence.

Growing up during one of the darkest times in history, Osamu Shimomura devoted his long and fruitful career to understanding how creatures emit light. He discovered green fluorescent protein (GFP), with which — decades later — biomedical researchers began to monitor the workings of proteins in living tissue, and to confirm the insertion of genes. For that discovery, he shared the Nobel Prize in Chemistry in 2008 with neurobiologist Martin Chalfie and the late Roger Tsien, a chemist.

Shimomura, who died in Nagasaki, Japan, on 19 October, was the first to show that a protein could contain the light-emitting apparatus within its own peptide chain, rather than interacting with a separate light-emitting compound. The significance of this discovery was that the gene encoding GFP could, in principle, be copied (or 'cloned') and used as a tool in other organisms. Others eventually took that step, but it would have been impossible without the exemplary patience of Shimomura, who spent years gathering enough material to extract, purify and determine the chemical structure of GFP.

Born on 27 August 1928 in the town of Fukuchiyama, at the height of Japanese expansionism, Shimomura was the son of an army captain whose frequent postings abroad disrupted his child's school education. Shimomura's grandmother instilled in him the samurai principles of honour and fortitude. In 1944, with the Pacific War turning against Japan, he and his fellow school students were mobilized to work in a munitions factory in Isahaya, about 25 kilometres from Nagasaki. On 9 August 1945, he was at work when a blinding flash, followed by a huge pressure wave, signalled the dropping of the atomic bomb on the nearby city. He walked home under a shower of black rain. He later wrote that his grandmother's quick action in putting him straight in the bath might have saved him from the effects of the radiation.

Without a high-school diploma, he despaired of finding a college place. Eventually, Nagasaki Pharmacy College admitted him in 1948. On graduation, he worked for four years as an assistant in practical classes. He devised research projects in his own time, and his professor obtained permission for him to do a year of advanced study. He joined the laboratory of organic chemist Yoshimasa Hirata at Nagoya University, and his lifelong fascination with bioluminescence began.
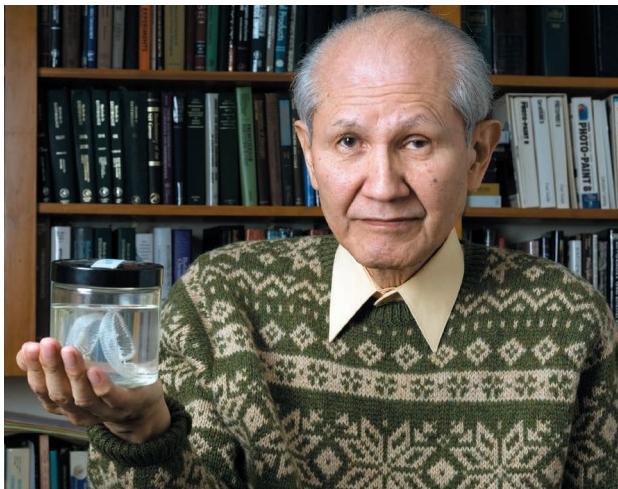
Hirata asked him to extract and purify a compound, luciferin, which enables the tiny marine crustacean *Cypridina* to glow in the dark. Hirata thought the results too uncertain for a PhD student, but because Shimomura was not registered for a degree, he allowed him to try. In just ten months, Shimomura made pure crystals of luciferin (O. Shimomura *et al. Bull. Chem. Soc. Japan* **30**, 929–933; 1957). "I learned that any difficult problem can be solved by great effort," he wrote in his Nobel biography.

The luciferin paper brought an invitation for Shimomura to join the bioluminescence lab of biologist Frank Johnson at Princeton University in New Jersey. Three weeks after marrying Akemi Okubo in August 1960, Shimomura sailed to the United States, his travel paid for by a Fulbright scholarship. Johnson asked him to work on the jellyfish *Aequorea*, which has a ring of organs around the edge of its umbrella that emit blue light. In July 1961, Johnson, Shimomura and his wife, and several assistants and students made a road trip across the United States to collect hundreds of jellyfish, scooping them out of Friday Harbor in Washington state, cutting off the rings and transporting them to Princeton for analysis.

In the face of scepticism from Johnson and others, Shimomura determined that the luminescent substance was a protein; he named it aequorin. He discovered almost at once that it was activated by calcium (later, aequorin became an essential reagent as a glowing marker of calcium release).

Shimomura, his family and his research colleagues spent 19 summers at Friday Harbor, collecting hundreds of thousands of jellyfish to obtain enough of the elusive material for a full structural analysis. Until a way of making genetically engineered aequorin became available in the 1990s, Shimomura freely shared his carefully harvested stocks with laboratories the world over.

It was in the process of purifying aequorin that Shimomura discovered small amounts of GFP, which fluoresces green when aequorin emits its blue light. It took him and his team until 1979 to accumulate enough to explore how the protein works. Shimomura described GFP's unprecedented incorporation of the light-emitting function within the protein chain (O. Shimomura *FEBS Lett.* **104**, 220–222; 1979), and then put GFP aside to work on wide-ranging studies of bioluminescence in other organisms.

In 1994, Chalfie's group reported the successful creation of bacteria and roundworms that could express GFP (M. Chalfie *et al. Science* **263**, 802–805; 1994). Soon afterwards, Tsien and his colleagues created GFPs of different colours (R. Heim *et al. Nature* **373**, 663–664; 1995). Others have extended the technique to vertebrates, with headlines about 'glow-in-the-dark monkeys' obscuring the method's great value in confirming the successful incorporation of genes from other organisms.

From 1982 until his retirement in 2001, Shimomura moved to the Woods Hole Marine Biological Laboratory in Massachusetts with Akemi, who had continued to work as his research assistant. After their retirement — a concept that was clearly difficult for them — they moved their laboratory into their home and continued to work. Shimomura's textbook *Bioluminescence: Chemical Principles and Methods* was released in 2006, and in 2017 he published an autobiography, *Luminous Pursuit: Jellyfish, GFP, and the Unforeseen Path to the Nobel Prize*. ∎

Georgina Ferry *is a science writer specializing in the history of the life sciences in Oxford, UK. Her many books include biographies of the crystallographers Dorothy Hodgkin and Max Perutz.*
e-mail: mgf@georginaferry.com

NEWS & VIEWS

Figure 1 | **Strong winds at Port Foster, Deception Island, Antarctica.**

CLIMATE SCIENCE

Warming linked to shifting winds

During the most recent ice age, abrupt changes in the Arctic climate were transmitted through the ocean to Antarctica. An atmospheric link between the two hemispheres has now been identified across the Antarctic continent. SEE LETTER P.681

NERILIE J. ABRAM

There are no precise analogues in Earth's past for the rapid warming that the world is now facing as a result of rising levels of greenhouse gases in the atmosphere. However, some periods of Earth's history can reveal valuable details of the way in which different parts of the climate system respond and interact[1]. Of particular interest are intervals of large and abrupt warming in the Arctic that occurred episodically during the most recent ice age (about 115,000–11,700 years ago). These climate shifts are known as Dansgaard–Oeschger events, and saw temperatures in Greenland jump by more than 10 °C in a matter of decades[2]. On page 681, Buizert et al.[3] report that these events altered the position of the westerly winds across the Southern Hemisphere — a finding that has implications for global ocean circulation and atmospheric carbon dioxide.

Past rapid warming events in the Northern Hemisphere give researchers a way of addressing fundamental questions in climate science. In particular, how do changes in the climate of one hemisphere affect that of the other? And how, and at what rate, do these changes propagate? The connection between hemispheres is important for determining how energy moves through the Earth system and alters the climate in different places around the globe.

The simplest model to explain the relationships seen between Greenland and Antarctic temperatures during Dansgaard–Oeschger events is the 'bipolar see-saw' movement of heat between the hemispheres through the global ocean[4]. In this model, the Greenland temperature jumps abruptly into its warm phase when the overturning circulation — the sinking of surface waters to the deep ocean — in the North Atlantic Ocean speeds up. This adjustment in ocean circulation concentrates heat in the Northern Hemisphere and causes

Antarctica to gradually cool. It takes about 200 years for the ocean changes in the North Atlantic to start affecting the Antarctic temperature[5]. This lag reflects the time that it takes for accumulated energy to penetrate north of the current that circles Antarctica and to begin to be absorbed into subsurface levels of the global ocean[6].

The opposite happens when the overturning circulation in the North Atlantic slows or stops, causing Greenland to shift quickly into a cold state similar to that associated with an ice age. It again takes about 200 years from when this rapid change occurs in the Arctic to when Antarctica begins to warm. The longer the cold state persists in Greenland, the more Antarctica warms through this see-saw mechanism of the deep ocean[7].

But this is not the full story. The atmosphere also provides a means by which climate signals propagate between the hemispheres, in a much faster way than through the ocean.

Abrupt warming events in the Arctic pull the meteorological equator — a band of tropical storm clouds that circle the globe near the Equator — farther north, and, along with it, the rainfall patterns associated with the Asian summer monsoon. In 2017, a reinterpretation of water-isotope signals in an Antarctic ice core identified a near-instantaneous response of atmospheric circulation to changes in Arctic climate that occurred in the most recent ice age, all the way south to West Antarctica[8]. However, whether this response occurred throughout the Southern Hemisphere, or was more localized, remained unclear.

Buizert and colleagues present the first Antarctic-wide evidence for a rapid atmospheric coupling of the position of the westerly winds around the whole of the Southern Ocean to past abrupt climate events in the Arctic (Fig. 1). Identifying these pervasive fluctuations in wind position, which happened on a decadal timescale tens of thousands of years ago, required the precise synchronization of ages for ice cores from across the Antarctic continent.

Ice-core ages from Greenland have been linked to those from Antarctica using the methane composition of bubbles in the exceedingly well-resolved ice core from the West Antarctic Ice Sheet Divide[5]. Atmospheric methane is quickly mixed across the hemispheres, and so can be considered as globally synchronous. Past fluctuations in methane abundance mimicked abrupt changes in Greenland temperature and therefore provide a way of precisely interrogating the timing of climate events between the Arctic and the Antarctic.

Buizert *et al.* took the next step in synchronizing the West Antarctic Ice Sheet Divide record with four other Antarctic ice cores by identifying characteristic sequences of volcanic eruptions preserved in the sulfate levels in Antarctic ice. Only then were the authors able to identify the superimposed oceanic and atmospheric signals that occurred across Antarctica in response to past rapid changes in Arctic climate.

The classic see-saw of heat between the hemispheres through the ocean can explain the delayed and gradual changes in Antarctic temperature that accompanied past abrupt shifts in Greenland temperature. But Buizert and co-workers' study suggests that superimposed on these slow ocean changes were an almost synchronous northward shift in the westerly winds circling Antarctica when Greenland moved into its warm phase — and, vice versa, a southward shift in these winds during cool Greenland events. This atmospheric response modulated the latitude in the Southern Ocean that formed the source of the moisture that fell as snow over Antarctica.

A one-to-one relationship has previously been identified between the duration of Greenland temperature events and the magnitude of the ensuing temperature response in Antarctica through the ocean mechanism[7]. Similarly, the

authors find that the atmospheric response seems to scale so that stronger Greenland events result in a larger climatic signal in Antarctica and the Southern Ocean. An atmospheric link tying changes in Arctic climate to the Antarctic has previously been hypothesized on the basis of climate-model responses in experiments designed to mimic aspects of Dansgaard–Oeschger events[6]. The current work provides the observational data to prove the existence of this link.

It is time to move beyond considering only the Atlantic Ocean and century-scale time lags when thinking about how the Arctic and the Antarctic are climatically connected[5]. Buizert and colleagues' identification of a rapid atmospheric link between climates at the poles has implications for our understanding of current climate change. Today, the Arctic is warming at about twice the rate of the global average; however, continent-scale warming of the Antarctic that is expected from climate simulations has not yet been clearly observed[9,10]. Changes in Antarctic sea ice are also not following expectations based on models[10]. Meanwhile, the westerly winds of the Southern Hemisphere have been shifting rapidly southwards, affecting water security in cities such as Perth in Australia and Cape Town in South Africa, and potentially having global consequences by altering the movement of heat and carbon dioxide between the atmosphere and the ocean[11].

Many challenges remain in accurately predicting how, and how quickly, the behaviour of Antarctica and the Southern Ocean will change in a warming climate. Nevertheless, the authors have provided a glimpse of the natural changes in behaviour — both rapid and slow — that occurred tens of thousands of years ago. These results provide a basis for progress in unravelling the current scientific mysteries of how the ocean and the atmosphere at the poles respond to rapid changes in climate. ∎

Nerilie J. Abram *is at the Research School of Earth Sciences and the Centre of Excellence for Climate Extremes, Australian National University, Canberra, ACT 2601, Australia.*
e-mail: nerilie.abram@anu.edu.au

1. Fischer, H. *et al. Nature Geosci.* **11**, 474–485 (2018).
2. Kindler, P. *et al. Clim. Past* **10**, 887–902 (2014).
3. Buizert, C. *et al. Nature* **563**, 681–685 (2018).
4. Stocker, T. F. & Johnsen, S. J. *Paleoceanography* **18**, 1087 (2003).
5. WAIS Divide Project Members. *Nature* **520**, 661–665 (2015).
6. Pedro, J. B. *et al. Quat. Sci. Rev.* **192**, 27–46 (2018).
7. EPICA Community Members. *Nature* **444**, 195–198 (2006).
8. Markle, B. R. *et al. Nature Geosci.* **10**, 36–40 (2017).
9. Abram, N. J. *et al. Nature* **536**, 411–418 (2016).
10. Jones, J. M. *et al. Nature Clim. Change* **6**, 917–926 (2016).
11. Toggweiler, J. R. & Russell, J. *Nature* **451**, 286–288 (2008).

ALZHEIMER'S DISEASE

A mosaic mutation mechanism in the brain

Variable brain–specific mutations have been observed in Alzheimer's disease. One mechanism underlying this mosaicism involves integration of variant gene copies back into the neuronal genome. SEE ARTICLE P.639

GUOLIANG CHAI & JOSEPH G. GLEESON

Genetic mutations can arise not only in fertilized eggs, affecting all cells of an organism, but also in a subset of an organism's cells[1–3]. The latter phenomenon, called mosaicism, is prevalent in the brain, and has been associated with several neurological disorders, including sporadic Alzheimer's disease, the most common form of the disease[1,3,4]. In 2015, it was found[5] that neurons from people with sporadic Alzheimer's contained more DNA and had more copies of the Alzheimer-related gene *amyloid-β precursor protein* (*APP*) than did neurons from people without the disease. However, the exact genomic changes underlying this mosaicism remained unresolved. Lee *et al.*[6] follow up on that work on page 639, providing a mechanism for increased

APP mosaicism in the brains of people with sporadic Alzheimer's disease. The study could alter our understanding of the roots of neurodegeneration.

First, Lee *et al.* set out to analyse *APP* variants in neuronal messenger RNA. In each experiment, the authors used mRNA from just 50 neurons from the brains of people with or without sporadic Alzheimer's, because averaging across large neuronal populations could mask variants present in only a few cells. The researchers' analysis revealed many *APP* mRNA variants. As expected, the variants lacked introns — non-protein-coding regions that are removed during gene transcription through a process called splicing, leaving only protein-coding exons. However, the variants were shorter than expected, and contained single-nucleotide mutations, inserted and

deleted exons, and larger deletions that led to the formation of new exon–exon junctions between missing multi-exon regions. Some of the mutations the authors observed have been previously implicated in familial Alzheimer's disease[7].

Lee and colleagues found the same short variants when they analysed genomic DNA from the neurons, suggesting that *APP*-variant mRNAs might be transcribed from matching genomic DNA sequences — named genomic complementary DNAs (gencDNAs) by the authors — that had become permanently embedded in the genomes of neurons. To further validate the existence of *APP* gencDNAs in neurons, the authors used two independent approaches: a technique called DNA *in situ* hybridization (DISH), in which fluorescent molecules were bound to gencDNA-specific exon–exon junctions in DNA; and sequencing of short sections of *APP* DNA. Both approaches confirmed the existence of gencDNA variants.

The researchers next investigated the extent of gencDNA diversity using DNA sequencing. In total, they identified 6,299 different *APP* gencDNA variants in 96,424 neurons from the brains of 5 people with sporadic Alzheimer's — approximately 10 times more than they found in the brains of people without the disease. In agreement, DISH also revealed substantially more gencDNAs in Alzheimer's neurons.

The authors demonstrated that *APP* gencDNAs were present in the neurons of a mouse model of Alzheimer's disease, but rarely in non-neuronal cells or neurons from control animals. Moreover, gencDNA variants accumulated with age. These findings are consistent with a role for *APP* gencDNA variants in the development of Alzheimer's. Indeed, the authors found that some *APP* mRNA variants are translated into proteins that are toxic to cells, further strengthening this possibility.

Finally, Lee and co-workers showed that gencDNAs could be generated in cells in culture, provided that two conditions were met. First, the cells' DNA had to contain breaks in its strands, and, second, the enzyme reverse transcriptase had to be active. This enzyme is responsible for a process called reverse transcription, in which matching DNA sequences are produced from mRNA. The data indicate that gencDNAs arise from reverse-transcribed mRNA intermediates, which are incorporated into the genome in a process that might be promoted by breaks in DNA (Fig. 1). In support of this idea, the authors detected reverse transcriptase activity in the human brain samples, and a previous study has shown the presence of DNA breaks in developing brains[8], whereas this phenomenon is rarely observed in other tissue types.

The incorporation of gencDNAs into the genome might share some mechanisms with retrotransposition — a process in which RNA
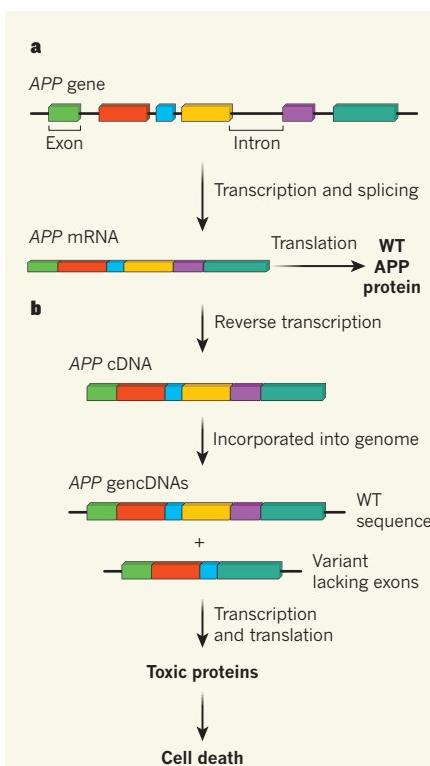
Figure 1 | Mosaic incorporation of *APP* variants into the neuronal genome. **a**, The gene *amyloid-β precursor protein* (*APP*) contains protein-coding exons (coloured blocks) and non-coding introns (this simplified schematic of the gene does not reflect the actual exon–intron composition). During transcription, introns are removed through a process called splicing to produce messenger RNA, which is translated to form the wild-type (WT) protein. **b**, Lee and colleagues[6] found that, in neurons in the human brain, *APP* mRNA undergoes a process called reverse transcription to produce a complementary DNA (cDNA). The cDNA can be reintegrated into the neuronal genome as a genomic cDNA (gencDNA). At some point in the process, mutations arise — perhaps when the cDNA is integrated into the genome, or at an earlier stage (not shown). This results in a range of gencDNA *APP* variants, some lacking one or more exons. Some gencDNA variants give rise to toxic proteins, leading to cell death. These processes might contribute to sporadic Alzheimer's disease.

transcribed from DNA sequences called transposable elements can reintegrate into new genomic regions to generate mosaicism[3]. But how gencDNAs become mutated from the original *APP* sequence remains unknown. Perhaps the mutations arise from mis-splicing of mRNA, or during genomic integration of gencDNAs.

Taken together, Lee and colleagues' work reveals the surprising existence of a phenomenon known as somatic gene recombination in the brain. This phenomenon, which has previously been reported only in antibody generation in immune cells[9], increases the diversity of proteins encoded by a given gene through DNA-shuffling mechanisms.

The study hints at a previously unanticipated mechanism in the development of Alzheimer's, and expands our understanding of the genesis of brain mosaicism. But whether accumulation of gencDNAs in neurons is a cause of or is caused by Alzheimer's disease remains to be proved.

The techniques used here could be applied to investigate whether gencDNA mechanisms are at work in other genes in other tissues; this could provide insights into diseases such as cancer or other degenerative disorders. However, it remains possible that gencDNA production is specific to *APP* or to neurons. The authors did not find gencDNA variants in another gene involved in Alzheimer's, *presenilin,* but nor did they rule out the possibility that gencDNAs could arise from other genes. Neurons have many features that might make them particularly vulnerable to gencDNAs: they are long-lived, have mostly stopped dividing, and have higher levels of reverse transcriptase activity and DNA-strand breaks than do non-neuronal cells[8].

It is also unclear whether the integration of *APP* gencDNAs into DNA is random or is biased towards certain genomic regions. The development of more-powerful sequencing techniques should help to answer this question.

Of course, there are many other avenues for further research. For instance, whether gencDNAs co-opt the retrotransposition and integration pathways used by transposable elements remains to be tested. The fact that gencDNAs are found in normal neurons suggests that they could have some benefits — this possibility should be examined. Finally, it will be interesting to test whether inhibitors of reverse transcriptase can prevent the accumulation of gencDNAs. Only when these avenues have been explored will we be able to build a complete picture of the remarkable phenomenon observed by Lee and colleagues. ∎

Guoliang Chai *and* **Joseph G. Gleeson** *are in the Department of Neurosciences, Howard Hughes Medical Institute, University of California, San Diego, La Jolla, California 92093, USA, and at the Rady Children's Institute for Genomic Medicine, Rady Children's Hospital, San Diego.*
e-mail: jogleeson@ucsd.edu; gchai@ucsd.edu

1. Biesecker, L. G. & Spinner, N. B. *Nature Rev. Genet.* **14**, 307–320 (2013).
2. Freed, D., Stevens, E. L. & Pevsner, J. *Genes* **5**, 1064–1094 (2014).
3. Richardson, S. R., Morell, S. & Faulkner, G. J. *Annu. Rev. Genet.* **48**, 1–27 (2014).
4. Beck, J. A. *et al. Hum. Mol. Genet.* **13**, 1219–1224 (2004).
5. Bushman, D. M. *et al. eLife* **4**, e05116 (2015).
6. Lee, M.-H. *et al. Nature* **563**, 639–645 (2018).
7. Murrell, J., Farlow, M., Ghetti, B. & Benson, M. D. *Science* **254**, 97–99 (1991).
8. Blaschke, A. J., Staley, K. & Chun, J. *Development* **122**, 1165–1174 (1996).
9. Tonegawa, S. *Nature* **302**, 575–581 (1983).

In Retrospect

A spotlight on bacterial mutations for 75 years

In the debate about how bacterial mutations arise, an experiment in 1943 showed that they can occur spontaneously and independently of a selection pressure. This study also popularized the use of maths–driven analysis of biological data.

MANOSHI S. DATTA & ROY KISHONY

Do bacteria acquire mutations randomly, or do mutations arise adaptively as a direct response to environmental pressures? This question has wide implications in areas ranging from evolution to the treatment of bacterial infections. In 1943, writing in *Genetics*, Luria and Delbrück[1] revealed, by a combination of experimental analysis and profound mathematical insight, that bacteria evolve through random mutations that arise independently of an environmental stress, and that occur even before bacteria encounter such selective conditions. Their study was a milestone in a debate about the nature and causes of bacterial evolution that is still ongoing. Moreover, this work has inspired the fields of microbial evolution and quantitative biology.

Luria and Delbrück worked at a time when scientists disagreed on the fundamental nature of bacterial evolution[2], despite tremendous advances in molecular biology and microbiology. For plants and animals, there was a general consensus that, consistent with Charles Darwin's theory of evolution, natural selection acted on mutations that arose randomly, regardless of their benefit to the organism. However, the unusual nature of bacterial genetics — such as the absence of sexual reproduction — sparked a vigorous debate about whether the principles that drive animal evolution also apply to bacteria (see go.nature.com/2brojqp). The main alternative hypothesis was Lamarckian evolution, named after the French biologist Jean-Baptiste Lamarck. In this model, the specific mutations that provide an advantage to an organism are acquired directly in response to the organism's environment[3].

For present-day microbiologists, this debate might seem strangely contrived — after all, if other organisms evolve in a manner consistent with the Darwinian principles of randomly occurring organismal variation that selection can act on, why should bacteria be an exception? Yet, it's worth having sympathy for our scientific predecessors. Even though we now accept that bacteria evolve through Darwinian mechanisms, 'quasi-Lamarckian' processes of bacterial evolution are still being discovered and debated[4–6].

Luria and Delbrück themselves encountered some difficulties when they entered the debate about how bacterial evolution occurs. To establish an approach to study mutations in bacteria, they allowed individual *Escherichia coli* cells to grow into large populations in individual test tubes, and added the cells from each of these tubes to Petri dishes containing agar coated with viruses known to kill the bacteria. Luria and Delbrück monitored the number of visible bacterial colonies on each of the plates. Each of these virus-resistant colonies arises from a cell and its descendants that had a mutation enabling the cells to survive the viral attack. Yet, for a simple experiment, their results were initially confusing: the number of colonies was highly variable between the different plates, a result that the authors initially attributed to an experimental error (see go.nature.com/2brojqp). But in a moment of clarity, Luria realized[7] that the high variability in the number of bacterial colonies might be an important clue, not an error.

Let's consider the experimental variance in the number of virus-resistant colonies per Petri dish expected under the process of either adaptive or random mutation. If mutations arise by an adaptive process, each bacterial cell would have a chance of acquiring a resistance mutation only on encountering the virus. Assuming each cell's chance of becoming resistant is small, the prediction would be that the number of virus-resistant colonies per Petri dish would vary according to a Poisson distribution (a standard probability distribution for random events, in which the standard deviation of the data equals the square root of the mean).

But, if evolution is driven by random mutations, mutations that confer viral resistance would arise during the growth of the bacterial population before viral exposure. In this case, the experimental variance in the number of virus-resistant bacterial colonies between different Petri dishes would be much higher than in the adaptive-mutation scenario, because the number of virus-resistant bacteria in a given test tube would depend on the random timing of when mutations occurred. A single virus-resistance mutation that occurred early in the growth of the bacterial population would result in a large number of virus-resistant bacterial descendants of the original mutated cell, whereas mutations that arose much later during the growth of the bacterial culture, just

50 Years Ago

After the wreck of the Torrey Canyon in March 1967, some 8,000 seabirds were taken to cleansing stations in Britain — but well under ten per cent … of these birds were rehabilitated and returned to the sea. Even this figure gives too optimistic a picture of the cleansing operation, for a large proportion of the so-called rehabilitated birds were recovered dead within a few days. Although exact figures are hard to come by, the Torrey Canyon episode revealed the complete inadequacy of the current methods of rehabilitating oiled birds … Legislation can never totally eliminate accidental pollution and it is estimated that even the much vaunted "load on top" system of washing tankers, although a great improvement on previous practice, produces pollution at a rate of 400,000 tons a year.
From *Nature* 30 November 1968

100 Years Ago

[U]ndoubtedly the war has been responsible for an enormous amount of destruction of capital; but when estimates are given … of the percentage of loss in Belgium, France, Italy, Serbia and other countries, it is not usually borne in mind that capital does not merely consist of gold and silver, of bricks and mortar … or even of railways, steamships and machinery … but of scientific knowledge … When, therefore, we compile estimates of the losses due to the war, let us not forget that our greatest asset, the vast store of knowledge that Science has gathered together for us … is still intact. It is a store that has slowly been accumulating ever since the beginning of the world — a store which enables man more and more to triumph over Nature, and one that for ever remains practically indestructible as the real permanent capital of the race, and by far its most precious heritage.
From *Nature* 28 November 1918

before viral encounter, would produce many fewer virus-resistant bacteria.

On the basis of this insight, Luria and Delbrück generated a statistical distribution (the Luria–Delbrück distribution) to describe the prevalence of virus-resistant bacterial mutants that would be expected if mutations arose randomly before the bacterial population came under selective pressure from the virus. Compared with a Poisson distribution expected for adaptive mutations, this Luria–Delbrück distribution has a long 'tail' at the end of the distribution pattern. In the context of the authors' experiments, this tail would correspond to Petri dishes that have a high number of bacterial colonies, corresponding to early mutational events that lead to a large number of mutant descendants.

The 1943 paper reported the results of the authors' experiments, termed fluctuation tests, that took this mathematical approach to analyse the number of virus-resistant colonies in *E. coli* populations. The authors' findings were consistent with mutations following a Luria–Delbrück distribution rather than a Poisson distribution, demonstrating that bacterial mutations arose randomly, and independently of an encounter with a virus.

Luria and Delbrück's work shaped subsequent studies of biology and evolution in many ways. Luria himself was reported as saying that their fluctuation test removed bacteria from "the last stronghold of Lamarckism" (see go.nature.com/2fbxujf). The fluctuation test is still a standard procedure for accurately measuring mutation rates in diverse systems, from bacteria[8] and yeast[9] to cancer cells[10]. Their study also popularized the use of *E. coli* and the viruses that attack it as a simple experimental model system for biology[11]. Beyond its direct impact in laboratories, the experiment became a textbook example of how mathematical thinking combined with simple experimentation can lead to profound biological insights[12]. For their contributions to bacterial and viral genetics, Luria and Delbrück won the Nobel Prize in Physiology or Medicine in 1969 (which they shared with the biologist Alfred Hershey).

Their insight into mutational processes also has implications in settings such as the clinic. In analogy to the original experiment, imagine a population of patients who have the same type of bacterial infection and who are being treated with the same antibiotic (the antibiotic replaces the virus as the selection pressure here). According to the random-mutation model, even if all else is equal among the patients, the number of antibiotic-resistant bacterial mutants initially present will vary highly between the patients, which could lead to markedly variable treatment outcomes. Because such high inherent variability in treatment efficiency reflects resistance mutations arising in a population before treatment, using DNA sequencing or other types of analysis to identify the presence and number

of antibiotic-resistant bacterial mutants before treatment could improve our ability to predict treatment outcome.

Did the Luria and Delbrück study really close the door on Lamarckism? As far as bacteria are concerned, the answer is much more complicated than the duo could probably ever have anticipated.

It is undeniable today that randomly occurring mutations and natural selection are central tenets of how bacterial evolution occurs[5]. However, scientists are uncovering and debating an increasing array of other evolutionary processes at work in bacteria, some of which are suspiciously Lamarckian in character[4–6]. For example, we now know that the genome-wide mutation rate, and even the mutation rates of specific genes, can be shaped by evolution and affected by the environment[13–15]. An even more striking example is bacterial adaptation through the CRISPR–Cas viral-defence system, in which bacteria can incorporate viral genetic material into their own genomes and use it, as an adaptive mechanism, to protect themselves and their descendants against current and subsequent viral attacks[16,17]. These quasi-Lamarkian mechanisms presumably evolved by random mutations and natural selection. They do not necessarily undermine the lessons learnt from Luria and Delbrück's work, but rather, show the power of evolution to sculpt living organisms in endlessly interesting ways.

It is intriguing to imagine an alternative scientific history that might have occurred if Luria and Delbrück had stumbled upon one of these quasi-Lamarkian mechanisms. The CRISPR–Cas defence mechanism is mainly repressed in the *E. coli* that they studied, but it is active in other bacterial species, such as *Streptococcus thermophilus*.

A fun challenge would be to repeat the Luria–Delbrück experiment under conditions that might favour the evolution of resistance by such adaptive mechanisms, for example by replacing *E. coli* with *S. thermophilus*. Would the distribution of the number of resistant mutants indicate random or adaptive mutations? What would Luria and Delbrück have concluded had they used a species that had the CRISPR–Cas system? The contingency of this historic choice underscores the fact that, like evolution, science perhaps also progresses both adaptively and randomly. ∎

Manoshi S. Datta *and* **Roy Kishony** *are in the Faculty of Biology, Technion–Israel Institute of Technology, Haifa 3200003, Israel. e-mails: mdatta@campus.technion.ac.il; rkishony@technion.ac.il*

1. Luria, S. E. & Delbrück, M. *Genetics* **28**, 491–511 (1943).
2. Zheng, Q. *Chance* **23** (2), 15–18 (2010).
3. Lenski, R. E. *PLoS Genet.* **13**, e1006668 (2017).
4. Koonin, E. V. & Wolf, Y. I. *Biol. Direct* **4**, 42 (2009).
5. Sniegowski, P. D. & Lenski, R. E. *Annu. Rev. Ecol. Syst.* **26**, 553–578 (1995).
6. Rosenberg, S. M. *Nature Rev. Genet.* **2**, 504–515 (2001).
7. Luria, S. E. *A Slot Machine, a Broken Test Tube: An Autobiography* (Harper & Row, 1984).
8. Bell, G. & MacLean, C. *Trends Microbiol.* **26**, 471–483 (2018).
9. Farlow, A. *et al. Genetics* **201**, 737–744 (2015).
10. Powers, S. & Pollack, R. E. *Nature Rev. Cancer* **16**, 266–270 (2016).
11. Cairns, J., Stent, G. S. & Watson, J. D. *Phage and the Origins of Molecular Biology* (Cold Spring Harb. Press, 2007).
12. Murray, A. *Genetics* **202**, 367–368 (2016).
13. Tenaillon, O., Denamur, E. & Matic, I. *Trends Microbiol.* **12**, 264–270 (2004).
14. Moxon, E. R., Rainey, P. B., Nowak, M. A & Lenski, R. E. *Curr. Biol.* **4**, 24–33 (1994).
15. Wielgoss, S. *et al. Proc. Natl Acad. Sci. USA* **110**, 222–227 (2013).
16. Marraffini, L. A. *Nature* **526**, 55–61 (2015).
17. Koonin, E. V. & Wolf, Y. I. *Biol. Direct* **11**, 9 (2016).

PALAEOANTHROPOLOGY

The not-so-dangerous lives of Neanderthals

Have Neanderthals gained an unfair reputation for having led highly violent lives? A comparison of skulls of Neanderthals and prehistoric humans in Eurasia reveals no evidence of higher levels of trauma in these hominins. SEE LETTER P.686

MARTA MIRAZÓN LAHR

Injuries are part of everyday life, from a scratch on the skin to a broken bone to a fatal trauma. Although many injuries are accidental, others can arise as a consequence of an individual's or a group's behaviour, activity or social norms — characteristics that tell us about societies and the inherent tensions and risks within and between different groups.

On page 686, Beier *et al.*[1] provide evidence that challenges the long-standing view[2] that Neanderthal populations experienced a level of traumatic injuries that was significantly higher than that of humans. The result calls into question claims[2,3] that the behaviour and technologies of Neanderthals exposed them to particularly high levels of risk and danger.

Reports of injuries and deaths are constantly in the news. As well as being drawn to read the

stories of individuals, such information is of interest because of what it tells us about our societies. However, to fully understand what might determine the current degree of violence and injuries, we also need to look back at the past and identify the causal underpinnings. But how far back should we look? Arguably, right back to the evolutionary origins of processes that shape behavioural, social and cognitive tendencies and abilities.

Anthropologists study skeletal remains to reconstruct aspects of ancient lives, building an 'osteobiography' that casts light on part of the life history of an individual. Skeletons preserve — in the form of holes, misshapen surfaces, bone misalignments and secondary fractures radiating out from a point of impact — a signature of the traumas that resulted in fractured, cut or perforated bones, even if the injuries subsequently healed[4,5].

Traumatic lesions have been frequently identified in Neanderthal fossils, particularly in the head (Fig. 1) and neck, leading to the view[2] that higher levels of skeletal injury occurred in Neanderthal populations than in human populations. However, this is not so, say Beier and colleagues. The authors assessed published descriptions of Neanderthal and modern human fossil skulls found in Eurasia from approximately 80,000 to 20,000 years ago. Comparing the number of injured and non-injured Neanderthal and human skulls, the authors report similar levels of head trauma in both groups.

The power of Beier and colleagues' analyses lies in their study design. Instead of comparing Neanderthal data with those of more-recent or living human populations, as previous studies have done[2,6], the authors based their comparisons on humans who not only shared aspects of their environment with Neanderthals, but whose fossil record also has a similar level of preservation. Beier et al. analysed data for 114 Neanderthal skulls and 90 human skulls. They gathered the data for 14 skull bones, and obtained information that ranged from 1 bone in poorly preserved fossils to data for all 14 bones per individual for well-preserved ones. In total, the authors recorded trauma incidence in 295 Neanderthal bones and 541 human bones. They also collected other information, such as the percentage of each of the 14 bones that was preserved for each individual, as well as details including sex, age at death and the fossil's geographic location.

Beier et al. ran two sets of statistical analyses — one based on the presence or absence of trauma in each of the skull bones, the other on individual fossil skulls as a whole — to test whether there were any statistically significant differences between the prevalence of trauma in the Neanderthal and human fossils. The authors also assessed whether trauma prevalence was linked to sex or age, taking into account fossil preservation, geographic location and possible interaction effects between the different variables.

Figure 1 | A Neanderthal skull. The Neanderthal fossil called Saint-Césaire 1. This fossil[9] shows signs of a healed bone injury[10] in the region indicated by the arrowheads. Beier et al.[1] assessed published analyses of ancient Neanderthal and human skulls, including that of Saint-Césaire 1. Contrary to the prevailing view[2] that Neanderthal existence was more violent than that of humans, the authors report that similar levels of trauma are present in Neanderthal and human fossils.

The two analyses gave similar results.

The more complete the fossils are, the more likely they are to have preserved evidence of injuries. This might seem obvious, but is an issue often ignored in such studies. Beier et al. offer a way to deal with this type of bias in the available material. Once the authors take into account the extent of fossil preservation, the predicted prevalence of trauma in Neanderthals and humans is almost the same.

Both Neanderthal and human males had a much greater incidence of trauma than did the females of their respective species. This pattern remains the same for humans today[7]. One final intriguing result is that, although traumatic injuries were present across all of the age ranges studied, Neanderthals that had trauma to the head were more likely to have died under the age of 30 than the humans were. The authors interpret this result as evidence that, compared with humans, Neanderthals either had more injuries when they were young

> *"Risk and danger were as much a part of the life of Neanderthals as they were of our own evolutionary past."*

or were more likely to have died after being injured.

Beier and colleagues' study does not invalidate previous estimates of trauma among Neanderthals. Instead, it provides a new framework for interpreting these data by showing that the level of Neanderthal trauma was not uniquely high relative to that of early humans in Eurasia. This implies that Neanderthal trauma does not require its own special explanations, and that risk and danger were as much a part of the life of Neanderthals as they were of our own evolutionary past. The result adds to growing evidence that Neanderthals had much in common with early human groups. However, the finding that Neanderthals might have experienced trauma at a younger age than humans, or that they had a greater risk of death after injury, is fascinating, and might be a key insight into why our species had such a demographic advantage over Neanderthals.

Is this the final word on the subject of Neanderthal trauma? The answer is no. Beier and colleagues assessed only skull trauma. What if Neanderthals accumulated more injuries to their bodies than did humans? There are data suggesting that this might be the case[8]. Furthermore, although the authors' analyses

CLAIRE ARTEMYZ

demonstrate the power of a well-designed study based on large samples, the data they used were recorded by many researchers and at varying levels of detail, raising the possibility of methodological biases.

Lastly, the causes of the injuries could provide some elusive insights into behaviour, activities or social norms in the past. From the shape, location and extent of traumatic injuries in skeletons, and characteristics such as the sharpness of fracture edges or the degree to which injuries had healed, it is sometimes possible to establish the most likely cause of a trauma — for example, whether the injury probably arose as a consequence of a hunting accident[3], interpersonal violence[10] or intergroup conflict[11]. Moreover, surviving severe

trauma might indicate that the injured person was cared for by members of their society[12]. Establishing the likelihood of each of these scenarios among Neanderthals and early modern humans will no doubt continue to challenge scientists for many years to come. ■

Marta Mirazón Lahr *is in the Department of Archaeology, Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Cambridge CB2 1QH, UK.*
e-mail: mbml1@cam.ac.uk

1. Beier, J., Anthes, N., Wahl, J. & Harvati, K. *Nature* **563**, 686–690 (2018).
2. Berger, T. D. & Trinkaus, E. *J. Archaeol. Sci.* **22**, 841–852 (1995).
3. Churchill, S. E. *Evol. Anthropol.* **11**, 185–186 (2002).
4. Redfern, R. C. *Injury and Trauma in Bioarchaeology. Interpreting Violence in Past Lives* (Cambridge Univ. Press, 2017).
5. Lovell, N. C. in *Biological Anthropology of the Human Skeleton* (eds Katzenberg, M. A. & Saunders, S. R.) 341–386 (Wiley, 2008).
6. Underdown, S. J. *Period. Biol.* **108**, 485–493 (2006).
7. Pinker, S. *The Better Angels of Our Nature: A History of Violence and Humanity* (Penguin, 2011).
8. Trinkaus, E. *J. Archaeol. Sci.* **39**, 3691–3693 (2012).
9. Lévêque, F. & Vandermeersch, B. *C. R. Acad. Sci.* **291**, 187–189 (1980).
10. Zollikofer, C. P. E., Ponce de León, M. S., Vandermeersch, B. & Lévêque, F. *Proc. Natl Acad. Sci. USA* **99**, 6444–6448 (2002).
11. Mirazón Lahr, M. et al. *Nature* **529**, 394–398 2016).
12. Spikins, P., Needham, A., Tilley L. & Hitchens, G. E. *World Arch.* https://doi.org/10.1080/00438243.2018.1433060 (2018).

This article was published online on 14 November 2018.

A glimpse into the heart of a quasar

Astronomical objects called quasars have been difficult to study because of the limited spatial resolution of observations. An approach has been developed that allows the structure and dynamics of quasars to be investigated. SEE LETTER P.657

ERIN KARA

When the astronomical object 3C 273 was detected[1], to most optical telescopes it looked just like a star in our Galaxy. But in 1963, astronomers discovered[2,3] that the object was shining from a distance of 750 megaparsecs (2.4 billion light years). Whatever this mystery object was, it was producing more radiation than a trillion stars, from a region no bigger than the Solar System. Objects such as 3C 273 are now known as quasars and are understood to be powered by hot gas and dust feeding into a supermassive black hole through a structure called an accretion disk. Fifty-five years after that remarkable discovery, 3C 273 is back in the limelight. On page 657, the GRAVITY Collaboration[4] reports observations of the spatially resolved rotation of hot gas in the quasar at distances much closer to the black hole than were previously possible.

A quasar can produce more energy than the entire galaxy in which it resides. Although the basic mechanism that powers a quasar is known, the anatomy of the supermassive black hole and its surroundings is not well understood. Where does the gas that feeds the black hole come from? And what effect does the resulting intense radiation have on the environment around the black hole? The findings of the GRAVITY Collaboration provide a way to answer these fundamental questions.

Determining the structure of a quasar is

difficult because the black hole is extremely small and far away from Earth, and therefore the gas orbiting close to the black hole cannot be directly imaged using telescopes. Instead, astronomers rely on the properties

of electromagnetic radiation coming from a single point to infer the structure and dynamics of the gas and dust around the black hole. Such properties include colour, time variability, polarization and phase — the offset of an electromagnetic wave from a given position.

For the past 30 years, our best understanding of gas in the vicinity of a quasar's black hole has come from a method called reverberation mapping, which uses echoes of light (analogous to those of sound) to map out regions near the black hole[5]. The accretion disk emits light in all directions, some of which is observed directly by telescopes, and some of which illuminates a region of surrounding gas, known to astronomers as the broad-line region. Optical-reverberation mapping measures how long it takes the broad-line region
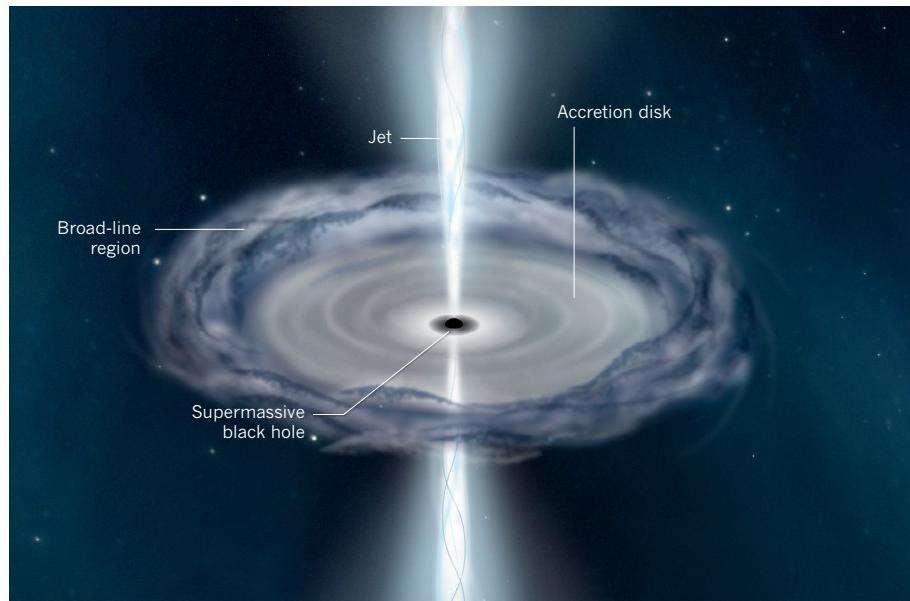
Figure 1 | Structure of the quasar 3C 273. Quasars are astronomical objects comprising a supermassive black hole surrounded by hot gas and dust. As this material is pulled towards the black hole through a structure known as an accretion disk, energy is released in the form of light and, in the case of the quasar 3C 273, as a beam of charged particles called a jet. The GRAVITY Collaboration[4] reports a technique that enables the rotation of gas in a part of 3C 273 known as the broad-line region to be spatially resolved. The researchers determine that this gas moves perpendicular to the jet and has the shape of a thick ring.

to respond to illumination from the accretion disk, which, in effect, measures the distance between the disk and the surrounding gas[6]. In a similar way to how bats use echolocation to map out a dark cave, astronomers measure light echoes to map out the hot gas around black holes.

The GRAVITY Collaboration has ushered in an alternative technique that spatially resolves the motion of such gas using the GRAVITY instrument in Chile[7]. This instrument is an interferometer that combines the light from four near-infrared telescopes that are 8 metres in diameter to produce a virtual 'super telescope' that is 130 m in diameter. Because the spatial resolution of a telescope depends on its size, the use of the GRAVITY instrument is a giant step in imaging capability. The collaboration measured the offset in phase between the direct emission of light from 3C 273 and the light from the broad-line region to spatially resolve the motion of this gas in a distant quasar for the first time.

The team observed a velocity gradient in the gas on size scales of 10 microarcseconds — an achievement that is comparable to seeing a coin on the Moon from Earth. The researchers found that the motion of this gas is perpendicular to the known large-scale jet (a beam of charged particles) projected from 3C 273 (Fig. 1). The results suggest that the gas is in the form of a thick ring with a radius of 0.12 parsecs, rotating around a black hole that has a mass 300 million times that of the Sun. These findings support previous estimates from reverberation mapping of 3C 273 that indicated a similar black-hole mass and gravitationally bound gas at a distance of 0.08–0.34 parsecs from the black hole[8,9].

For astronomers, the excitement about the current work is not because the results have fundamentally changed our understanding of quasars, but rather because this impressive technological advance enables an independent cross-check of optical-reverberation mapping — the most widely used method for determining the structure of gas around supermassive black holes. Optical reverberation has been measured in roughly 60 quasars[10], and the inferred properties of the gas strongly correlate with the luminosity of the quasar and the mass of the central black hole.

These correlations have been applied to large samples that comprise thousands of quasars. They have thereby informed our understanding of far-reaching aspects of astronomy, from the co-evolution of black holes and galaxies over cosmic time to the rate at which the expansion of the Universe is accelerating. Having an independent cross-check from spatially resolved interferometric observations, as reported by the GRAVITY Collaboration, is valuable for confirming several key findings in astrophysics that rely on the robustness of reverberation-mapping results.

It is important to keep in mind that the results presented in the paper are based on one particular quasar. The GRAVITY Collaboration observed 3C 273 because it was the best target for optical interferometry. However, the quasar is by no means the best target for reverberation mapping, which makes it difficult to compare the results from these two methods critically.

Going forward, the GRAVITY instrument should be capable of spatially resolving the dynamics and orientations of the broad-line region in about ten other quasars[11]. To best corroborate or dispute sizes and structures inferred from reverberation mapping, coordinated campaigns on the same quasars using two independent techniques must be carried out. The GRAVITY instrument is at the beginning of its scientific operations, and these early technical achievements bode well for future investigations that peer deeper into the hearts of quasars. ■

Erin Kara *is in the Department of Astronomy, University of Maryland, Maryland 20742, USA.* e-mail: ekara@astro.umd.edu

1. Edge, D. O., Shakeshaft, J. R., McAdam, W. B., Baldwin, J. E. & Archer, S. *Mem. R. Astron. Soc.* **68**, 37–60 (1959).
2. Hazard, C., Mackey, M. B. & Shimmins, A. J. *Nature* **197**, 1037–1039 (1963).
3. Schmidt, M. *Nature* **197**, 1040 (1963).
4. GRAVITY Collaboration. *Nature* **563**, 657–660 (2018).
5. Blandford, R. D. & McKee, C. F. *Astrophys. J.* **255**, 419–439 (1982).
6. Peterson, B. M. *Publ. Astron. Soc. Pacif.* **105**, 247–268 (1993).
7. GRAVITY Collaboration. *Astron. Astrophys.* **602**, A94 (2017).
8. Kaspi, S. *et al. Astrophys. J.* **533**, 631–649 (2000).
9. Zu, Y., Kochanek, C. S. & Peterson, B. M. *Astrophys. J.* **735**, 80 (2011).
10. Bentz, M. C. & Katz, S. *Publ. Astron. Soc. Pacif.* **127**, 67–73 (2015).
11. Rakshit, S., Petrov, R. G., Meilland, A. & Hönig, F. *Mon. Not. R. Astron. Soc.* **447**, 2420–2436 (2015).

MICROBIOLOGY

A bacterium's enemy isn't your friend

The bacterium *Staphylococcus aureus* is a leading cause of hard-to-treat human infections. It now seems that, if the bacterium is infected by a virus, a viral enzyme helps the microbe to evade detection by the immune system. SEE LETTER P.705

MICHAEL S. GILMORE & ONA K. MILLER

Microorganisms thrive on our body's surfaces. The species present are not just a random assembly; rather, they are a community of organisms that are particularly well adapted to the local conditions of temperature, moisture, nutrient availability and host defences[1]. *Staphylococcus aureus* is one of our most common bacterial residents. It usually lives in nasal, respiratory and reproductive tissues without causing disease, yet, unlike many other resident bacteria, *S. aureus* has the capacity to give rise to a potentially deadly infection[2].

During the past 50 years[2], the resistance of *S. aureus* to antibiotics has become an increasing problem, and strains of the bacterium termed methicillin-resistant *S. aureus* (MRSA), which are resistant to treatment with the antibiotic methicillin and other methylated penicillin-based antibiotics, cause both hospital- and community-acquired infections around the globe. On page 705, Gerlach *et al.*[3] describe a previously unknown mechanism whereby viruses influence whether MRSA is recognized by the immune system, shedding light on a process that might tip the balance in determining whether this bacterium will be harmless or disease-causing.

Staphylococcus aureus belongs to the Gram-positive group of bacteria, and has been described as existing on the borderline between being a normal human microbial resident and a disease-causing organism[4]. This bacterium seems to have the capacity to probe for signs of host weakness, such as reduced immune defences caused by disease. When this is detected, the bacterium can increase its population to a level that can cause the death of the host[5]. Factors that regulate host–microbial interactions are complex, and in addition to host defences, such interactions can be influenced by the presence or absence of other bacteria[6]. Gerlach and colleagues report that viruses can also be part of the mix that influences host–microbial interactions in the context of MRSA.

In Gram-positive bacteria, the cell wall contains polymers known as wall teichoic acids (WTA), which are made up of ribitol phosphate or glycerol phosphate molecules and can constitute up to half of the cell-wall mass[6]. Unlike the other main cell-wall component, peptidoglycan, which forms a porous and comparatively insoluble meshwork, WTA form a highly hydrated, gel-like material that fills much of the space between peptidoglycan strands. WTA provide a soluble matrix through which all substances pass before reaching the bacterial cell membrane, and therefore affect bacterial access
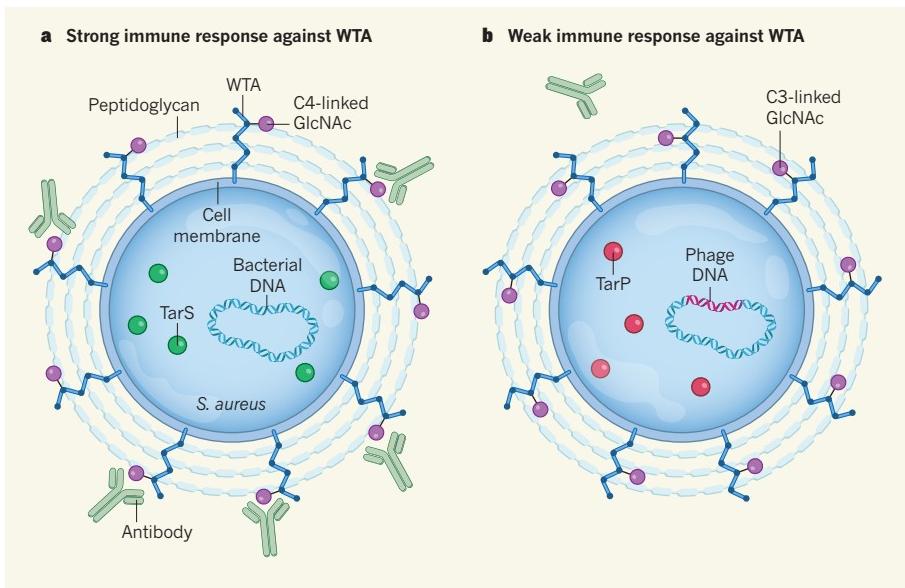
Figure 1 | Viral infection of a bacterium can alter the host's immune response to the microbe. a, The bacterium *Staphylococcus aureus* is a resident of the human body. Its outer surface is coated with layers of the polymer peptidoglycan crosslinked to wall teichoic acids (WTA) — polymers of ribitol phosphate molecules. The bacterial enzyme TarS modifies WTA, generating a form that has the molecule GlcNAc attached at carbon atoms in the C4 position in the ribitol. Human antibodies against *S. aureus* often target WTA. **b**, Gerlach *et al.*[3] report that some antibiotic-resistant strains of *S. aureus,* which are associated with difficult-to-treat infections, have been infected with a virus called a phage. The phage DNA encodes an enzyme called TarP that attaches GlcNAc to WTA at carbon atoms in the C3 position of ribitol rather than in the normal C4 position. In studies of mice and of human cells, the authors find that these TarP-modified WTA trigger an immune response that is weaker than the response against TarS-modified WTA.

to ions, nutrients, proteins and antibiotics[7]. In *S. aureus*, WTA are composed of units of D-ribitol phosphate, which are crosslinked to the peptidoglycan (Fig. 1). WTA function is tuned by attachments of the amino acid D-alanine and of *N*-acetylglucosamine (GlcNAc)[7] molecules to the ribitol-phosphate polymer.

Gerlach and colleagues decided to investigate whether bacterial evasion of immune-system defences might be one of the reasons that MRSA strains can reach high enough bacterial numbers to cause disease. The authors studied the genome sequences of MRSA strains to identify genes encoding enzymes that modify WTA. This revealed that some MRSA strains encode an enzyme called TarP that catalyses the addition of GlcNAc to D-ribitol phosphate at a particular carbon atom (known as C3) in the ribitol. Normally, GlcNAc is added at a different position, the C4 carbon, by the action of a related enzyme called TarS.

Surprisingly, the TarP-encoding sequence is of viral origin, and is found in *S. aureus* as a result of infection by a bacterial virus called a phage. TarP is dominant over its bacterial counterpart, TarS — that is, if both enzymes are present, the GlcNAc linkage is made on the C3 carbon of ribitol, rather than on the C4 carbon. *S. aureus* is normally held in check because the immune system has the ability to detect it. However, the authors found that, in mice, the form of WTA made by TarP action is less likely to trigger an immune response than

is the form of WTA generated by TarS.

This virus-mediated change to the *S. aureus* cell wall reported by Gerlach and colleagues is important for two reasons. First, it highlights the fact that a fragile truce between host and resident microbe can be affected by the intervention of a third party with its own vested interests. Second, at a time that some[8] have called the beginning of a 'post-antibiotic era' — given the rise in antibiotic-resistant bacteria and the limited development of new antibiotics reaching the clinic — there is a pressing need to develop new strategies to manage infection.

> "A fragile truce between host and resident microbe can be affected by the intervention of a third party."

We are now at the dawn of a clinical era in which the goal will be to precisely manage human and microbial interactions to promote health and limit disease. Antibiotics will continue to have a key role, as undoubtedly will other approaches, including the replacement of a person's gut microbes using techniques such as faecal transplants, or the use of phage-mediated elimination of undesirable microbes. Determining the best approach will be helped by the development of new diagnostic tools and a clearer understanding of the nature of human and microbial interactions. If deciding whether to take an approach based on a vaccine or possibly using phage treatments

in the future, key considerations will include knowing how a bacterium's susceptibility to phage infection varies, and determining whether the presence of phage DNA in a bacterial genome affects the dynamics between human cells and the microbes that colonize the body.

We do not yet know whether the phage-mediated alteration of WTA described by Gerlach and colleagues affects where the bacteria reside on the body or the number of bacterial cells present. We also lack a clear understanding of whether the antistaphylococcal WTA-targeting antibodies that most people have, and which do not seem to be protective against infection in immune-deficient individuals, are a 'distraction' imposed by the presence of *S. aureus*. This distraction would keep the immune system busy generating antibodies that end up in ineffective locations such as the bloodstream and do not eliminate the microbe. Alternatively, this low-level immune warfare could represent a stalemate between the host and its resident bacteria.

It is clear that phage-encoded TarP changes the immune reactivity of *S. aureus*. In a model system of human immune cells grown *in vitro*, the authors found that *S. aureus* strains encoding TarP were cleared from the system less effectively than were *S. aureus* strains that lacked TarP. Similar phage-mediated changes in a bacterial cell surface that alter antibody recognition of the microbe have been reported[9] for the disease-causing Gram-negative bacterium *Shigella flexneri*.

Gerlach and colleagues' work, as well as that of others in this area, demonstrates that the balance between host and microbes is a dynamic one. The discovery that phages can have a role in tipping the delicate balance between *S. aureus* colonization and infection might one day affect the choice of approaches for treating MRSA infections. ■

Michael S. Gilmore and **Ona K. Miller** *are in the Departments of Ophthalmology and of Microbiology and Immunobiology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, USA. e-mails: michael_gilmore@meei.harvard.edu; ona_miller@meei.harvard.edu*

1. Proctor, D. M. & Relman, D. A. *Cell Host Microbe* **21**, 421–432 (2017).
2. Lee, A. S. *et al. Nature Rev. Dis. Primers* **4**, 18033 (2018).
3. Gerlach, D. *et al. Nature* **563**, 705–709 (2018).
4. Camargo, I. L. B. C. & Gilmore, M. S. *J. Bacteriol.* **190**, 2253–2256 (2008).
5. Guerra, F. E., Borgogna, T. R., Patel, D. M., Sward, E. W. & Voyich, J. M. *Front. Cell Infect. Microbiol.* **7**, 286 (2017).
6. Boldock, E. *et al. Nature Microbiol.* **3**, 881–890 (2018).
7. Neuhaus, F. C. & Badilly, J. *Microbiol. Mol. Biol. Rev.* **67**, 686–723 (2003).
8. Zucca, M. & Savoia, D. *Int. J. Biomed. Sci.* **6**, 77–86 (2010).
9. Mavris, M., Manning, P. A. & Morona, R. *Mol. Microbiol.* **26**, 939–950 (1997).

ARTICLE

Somatic *APP* gene recombination in Alzheimer's disease and normal neurons

Ming-Hsiang Lee[1], Benjamin Siddoway[1,3], Gwendolyn E. Kaeser[1,2,3], Igor Segota[1,3], Richard Rivera[1], William J. Romanow[1], Christine S. Liu[1,2], Chris Park[1,2], Grace Kennedy[1], Tao Long[1] & Jerold Chun[1]*

The diversity and complexity of the human brain are widely assumed to be encoded within a constant genome. Somatic gene recombination, which changes germline DNA sequences to increase molecular diversity, could theoretically alter this code but has not been documented in the brain, to our knowledge. Here we describe recombination of the Alzheimer's disease–related gene *APP*, which encodes amyloid precursor protein, in human neurons, occurring mosaically as thousands of variant 'genomic cDNAs' (gencDNAs). gencDNAs lacked introns and ranged from full-length cDNA copies of expressed, brain-specific RNA splice variants to myriad smaller forms that contained intra-exonic junctions, insertions, deletions, and/or single nucleotide variations. DNA in situ hybridization identified gencDNAs within single neurons that were distinct from wild-type loci and absent from non-neuronal cells. Mechanistic studies supported neuronal 'retro-insertion' of RNA to produce gencDNAs; this process involved transcription, DNA breaks, reverse transcriptase activity, and age. Neurons from individuals with sporadic Alzheimer's disease showed increased gencDNA diversity, including eleven mutations known to be associated with familial Alzheimer's disease that were absent from healthy neurons. Neuronal gene recombination may allow 'recording' of neural activity for selective 'playback' of preferred gene variants whose expression bypasses splicing; this has implications for cellular diversity, learning and memory, plasticity, and diseases of the human brain.

The diversity of neuronal form and function is intrinsic to the human brain, but its basis remains largely unknown. Early speculations involved gene recombination[1], analogous to the mechanism of antibody diversification that was later identified[2], but this has not been described in the brain[3,4]. Nevertheless, later identification of genomic mosaicism[5], which arises somatically to produce brain cells with distinct if seemingly random genomic changes, suggested genome dynamism that might include gene recombination. Genomic mosaicism was first identified in neural progenitor cells and neurons as aneuploidies and DNA content variation, both representing large copy number variations (CNVs)[6–8]. Randomly distributed, smaller megabase-scale CNVs, LINE1 repeat elements, and single nucleotide variations (SNVs) were subsequently identified. Genomic mosaicism can influence cell survival and gene transcription, but somatic gene recombination of specific genes has not been reported[5,9].

A candidate gene for neuronal recombination is *APP*, which shows mosaic CNVs in normal human brains. These CNVs are increased in sporadic Alzheimer's disease (SAD)[10], the most common form of Alzheimer's disease. *APP* is central to the amyloid hypothesis wherein APP is cleaved by secretases to form toxic amyloid-β (Aβ) peptides and plaques, causing Alzheimer's disease[11]. Constitutive *APP* mutations and duplications are believed to cause rare forms of familial Alzheimer's disease (FAD) and Alzheimer's disease neuropathology in Down syndrome (trisomy 21 with 3 *APP* copies), supporting the idea that they have a pathogenic role when present mosaically in SAD[12–14]. We previously identified mosaic, neuronal *APP* CNVs that showed heterogeneous signals that might be explained by gene recombination[10]. However, interrogation of *APP* genomic loci (about 0.3 Mb) using low-depth, short-read single-cell sequencing capable of detecting CNVs produced negative results that were complicated by resolution limitations[5,15]. We therefore developed an alternative strategy focused on *APP* in small

cell populations, using nine distinct methodologies (Extended Data Table 1).

Novel *APP* RNA variants in neurons

We postulated that genomic sequence alterations in *APP*, existing mosaically, could be detected in RNA through transcriptional amplification. Assessments were focused on small populations of nuclei rather than bulk samples that are dominated by annotated species (Extended Data Fig. 1a) to detect mosaic alterations. The workflow (Fig. 1a) commenced with fluorescence-activated nuclear sorting (FANS)[16] to isolate neuronal nuclei from prefrontal cerebral cortices from both control individuals and those with verified SAD, which were run in parallel (Extended Data Table 2). Groups of 50 NeuN-positive neuronal nuclei were isolated and processed for PCR with reverse transcription (RT–PCR; Fig. 1a) and downstream analysis. RT–PCR using validated primers on exon 1 and exon 18 (Supplementary Table 1), which can amplify full-length *APP* cDNA (*APP*-770, NM_000484.3), detected the expected splice variants *APP*-751 (NM_201413.2) and *APP*-695 (NM_201414.2)[17] (Extended Data Fig. 1b). However, multiple unexpected bands of varied sizes were also identified (Fig. 1b). The RT–PCR products were Southern blotted with ³²P-labelled APP cDNA probes (Fig. 1c), and positive bands were cloned and Sanger sequenced. The new bands yielded *APP* cDNA sequence variants unlike any previously reported, characterized by loss of central exons with proximal and distal exons linked by intra-exonic junctions (IEJs) (Fig. 1d,e). Twelve novel RNA variant sequences with unique IEJs were identified in neurons (Fig. 1e) and non-neurons displayed no variants (Extended Data Fig. 1c). IEJs were independently observed in five oligo-dT-primed cDNA libraries; three from sorted neuronal nuclei from individuals with SAD (Extended Data Fig. 1d) and two from commercially
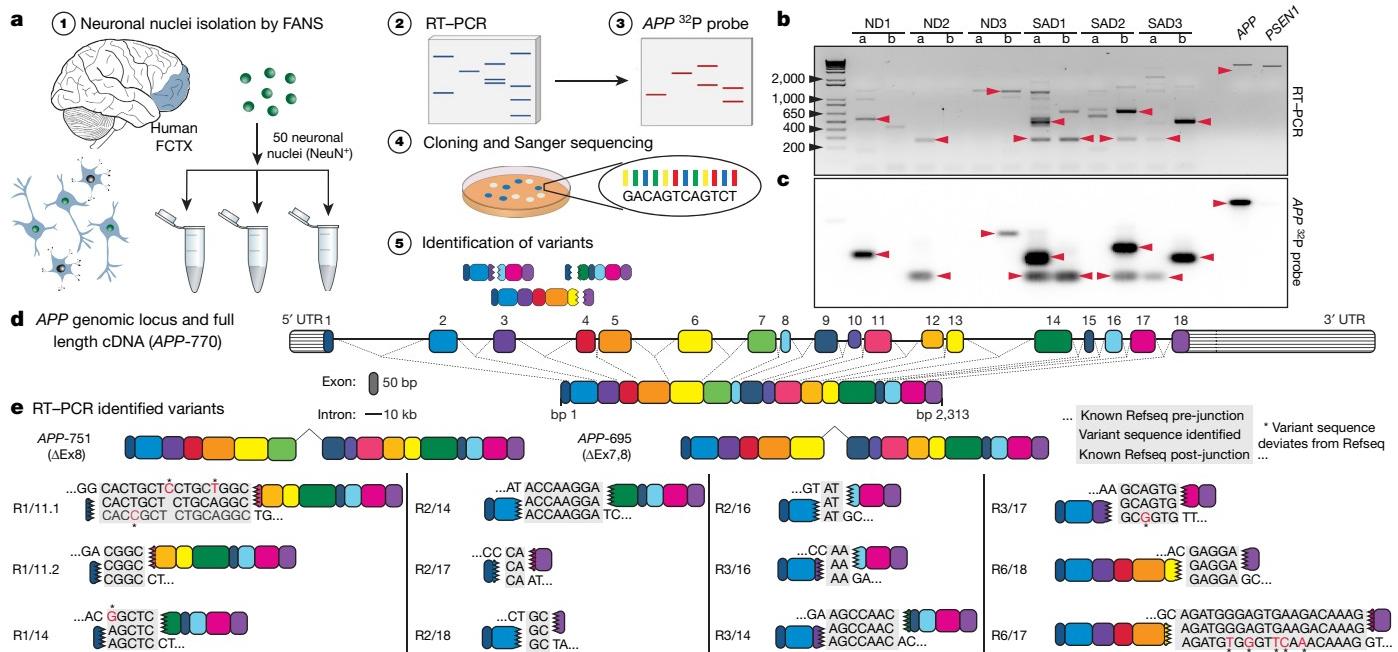
Fig. 1 | Identification of novel APP RNA variants from small populations of neurons. a, Fifty neuronal nuclei were sorted from human prefrontal cortices (FCTX) (1) and used for RT–PCR (2). The resulting RT–PCR products were screened by Southern blotting with ^{32}P-labelled APP cDNA probes (3). Bands with positive signals from duplicate gels were cloned and sequenced (4), and variants were identified (5). **b**, Electrophoresis of RT–PCR products from the brains of three non-diseased (ND) individuals and three patients with SAD, with two populations each (a and b). APP and PSEN1 plasmids were run as positive and negative controls for Southern blotting. **c**, Southern blot of RT–PCR products. Arrowheads indicate examples of bands from **b** that were cloned and Sanger sequenced. **d**, Structure of human APP genomic locus and spliced APP-770 full-length cDNA drawn to scale; the colour scheme remains consistent throughout all figures. **e**, APP RNA variants identified by RT–PCR. The sequences of homology regions forming IEJs are shown. Variant sequences deviating from Refseq are shown in red with asterisks. R, RNA identified; #/#, exon–exon junction; .#, for multiple unique junctions.

produced long-read RNA-seq data sets from whole brain and temporal lobe of patients with SAD (Extended Data Fig. 1e). Five variants retained coding potential and seven contained premature stop codons (Extended Data Table 3, Supplementary Table 2). One prevalent form was characterized by an IEJ between the 24th nucleotide of exon 3 and the 45th nucleotide of exon 16 (Fig. 1e, R3/16). Detection of R3/16 by RNA in situ hybridization (RISH) on SAD brain sections indicated the cytoplasmic presence of variants (Extended Data Fig. 1f). Notably, sequence complementarity of joined exons was found in all 12 IEJs, ranging in overlap from 2 to 20 nucleotides (Fig. 1e, Extended Data Table 3, Supplementary Table 2). Amplification for a second gene related to Alzheimer's disease, PSEN1, did not identify variants (Extended Data Fig. 1g).

gencDNA sequences in neuronal genomes

The existence of previously unidentified RNA variants raised the question of whether this transcriptional heterogeneity originated from mosaic variation in DNA. We carried out high-stringency amplification, using the APP primers previously used for RNA and cDNA analyses, on RNase-treated DNA extracted from sets of 20 neuronal nuclei from both healthy brains and those with SAD (Fig. 2a). PCR of the wild-type APP genomic locus was not possible because of its length (about 300 kb) (Fig. 1d). However, PCR on genomic DNA generated similar-sized bands to novel RNA variants (Fig. 2b, Extended Data Fig. 2a). Sanger sequencing revealed multiple gencDNAs and seven of eight were identical to those identified in RNA (Fig. 2c). We validated the presence of APP gencDNAs in neurons using multiple, distinct primer sets (Extended Data Fig. 2b, c). We did not detect gencDNAs in DNA isolated from human lung fibroblasts (IMR-90), human embryonic kidney cells (HEK-293), or non-neuronal nuclei from the brains of individuals with or without SAD (Extended Data Fig. 2d, e). Amplification of PSEN1 did not produce products from genomic DNA (Fig. 2b, Extended Data Fig. 2a).

gencDNA detection by non-PCR methods

To validate the presence of APP gencDNA junctions within single neuronal genomes without polymerase-based amplification, we developed DNA in situ hybridization (DISH). Our method extensively modified the sample preparation and hybridization protocols (see Methods) of a commercial RISH product, BaseScope (ACD), to recognize genomic sequences. BaseScope technology uses paired ISH probes to eliminate hybridization artefacts and can detect specific junctions. Two DISH probes were extensively used (Extended Data Table 4): one that recognized a common gencDNA sequence via the exon 16–exon 17 junction (DISH$_{16/17}$), which spans the Aβ coding region of APP; and one that recognized the newly identified IEJ formed between exons 3 and 16 (DISH$_{3/16}$). Bound probes were visualized as red dots with varying diameters. All probes passed multiple specificity requirements involving positive and negative controls. Sense and antisense DISH probes produced similar results in RNase-treated neuronal nuclei from individuals with SAD (Fig. 2d–i). By comparison, RNA signals were detected only using the antisense probes (Extended Data Fig. 1f); therefore, sense probes were used in all subsequent DISH analyses. Critically, DISH signals were eliminated by destruction of the target sequence by specific (but not off-target) restriction enzyme digestion (Fig. 2j–m, Extended Data Fig. 2f). In addition, no DISH signal was detected on cells infected with retroviruses containing wild-type human genomic APP sequences lacking target sequences (Extended Data Fig. 2g, h). Notably, double labelling with dual DISH probes recognizing the intron 2–exon 3 wild-type genomic sequence combined with DISH$_{3/16}$ or DISH$_{16/17}$ demonstrated that APP gencDNAs did not usually co-localize with the wild-type locus (Fig. 2n). Thus, DISH detected specific APP gencDNA junctions within genomic DNA without polymerase-dependent amplification, revealing multiple loci distinct from germline APP alleles.

A completely independent approach also identified APP gencD-NAs without primary PCR amplification by using a custom Agilent
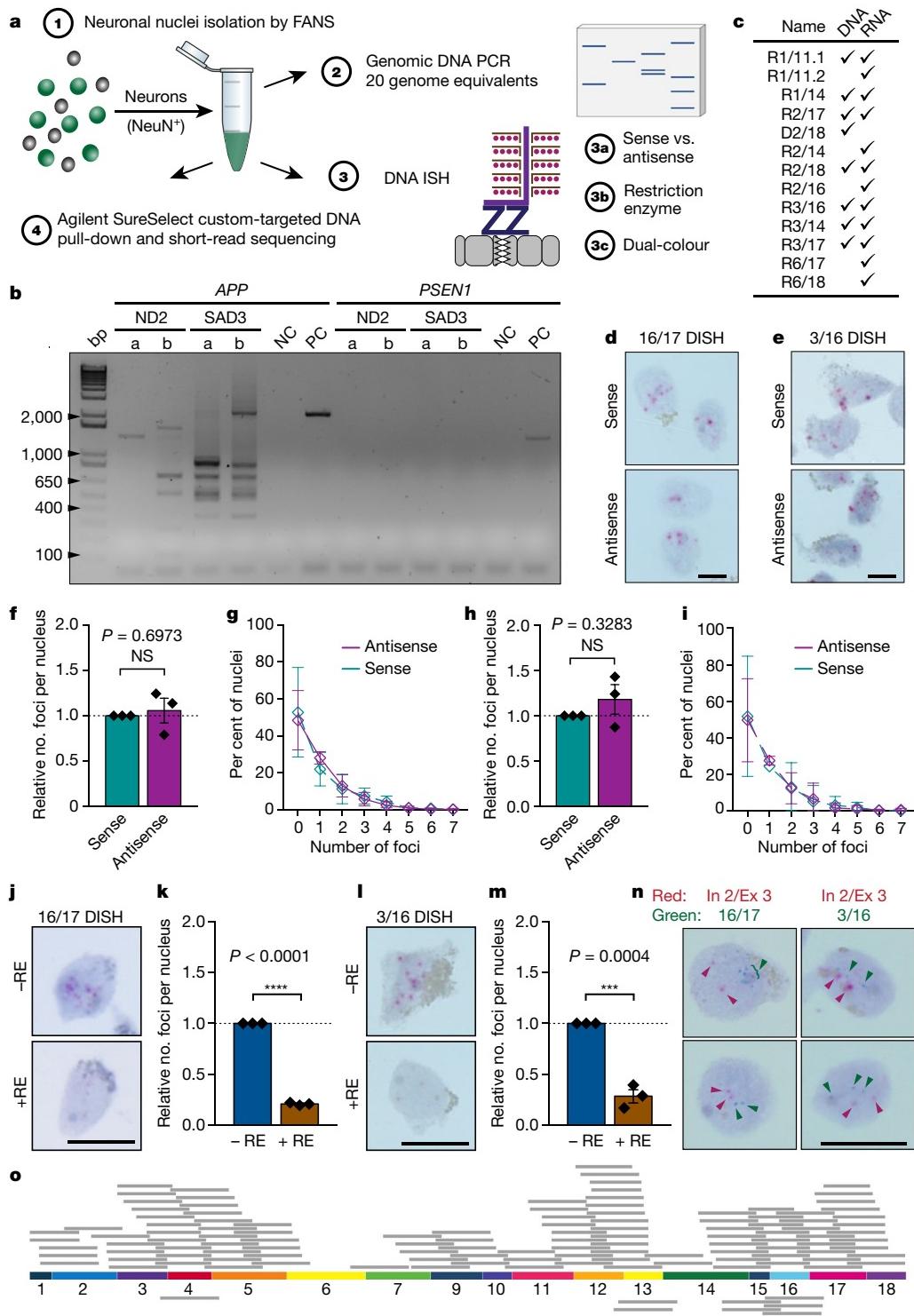
Fig. 2 | *APP* gencDNAs identified by DNA polymerase-dependent and -independent methods. a, FANS-isolated neuronal nuclei from human prefrontal cortices (1) were used for genomic DNA PCR (2), DISH (3), and custom target enrichment followed by deep sequencing (4). **b**, Electrophoresis of genomic DNA PCR products with *APP* and *PSEN1* primer sets, from neurons from normal brains or brains from individuals with SAD with two replicates (a and b). Non-template control (NC) and positive control (PC) with indicated plasmids are shown. **c**, Cloning and Sanger sequencing revealed multiple gencDNA sequences. **d–i**, DISH was performed with sense and antisense probes targeting the exonic 16/17 junction (**d, f, g**; sense $n = 339$ and antisense $n = 335$), and the intra-exonic 3/16 junction (**e, h, i**; sense $n = 490$, antisense $n = 484$) on neuronal nuclei from individuals with SAD. **f–i**, Relative average number of foci

(**f, h**; points represent average of independent experiments) and frequency distributions (**g, i**) showed no significant differences (unpaired, two-tailed Student's *t*-test). **j–m**, Restriction enzyme (RE) digestion using MluCI (**j, k**) and PstI + MslI (**l, m**) to eliminate 16/17 (−RE $n = 349$, +RE $n = 440$) and 3/16 (−RE $n = 367$, +RE $n = 340$) target sequences, respectively. Statistical significance on all bar graphs was determined using unpaired, two-tailed Student's *t*-test. **n**, Dual DISH with intron 2/exon 3 (red) genomic locus and 16/17 or 3/16 probes (green). **o**, Schematic of *APP* cDNA and genomic exon–exon junctions identified by Agilent SureSelect enrichment of the *APP* locus and Illumina sequencing; reads below span two exon–exon junctions. NS, not significant. Error bars show s.e.m. Scale bars, 10 μm.
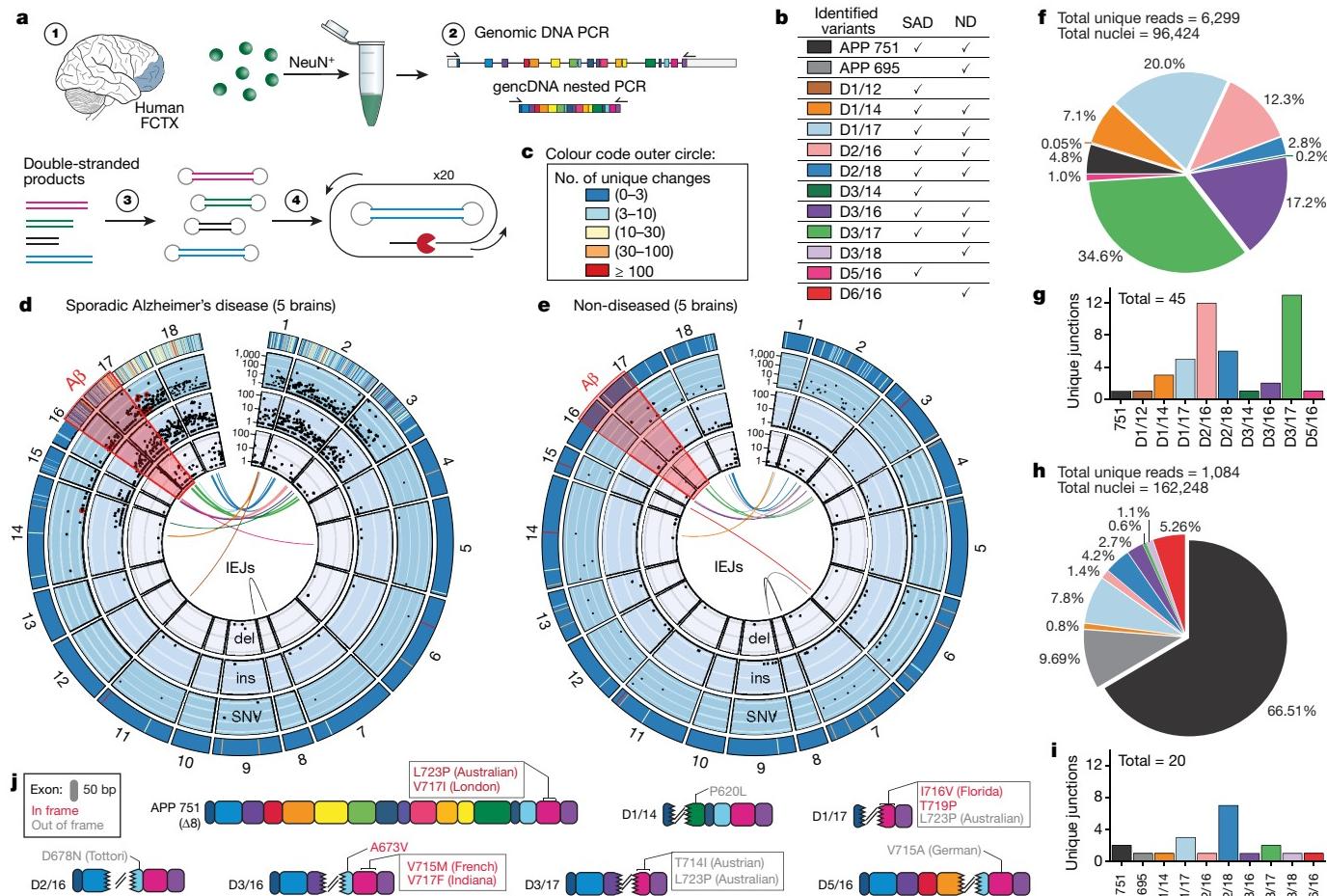
Fig. 3 | Thousands of unique gencDNAs identified by SMRT sequencing of neurons from non-diseased brains and brains of individuals with SAD. a, Neuronal nuclei from prefrontal cortices of individuals with or without SAD were sorted (1) and used for genomic DNA PCR (2). Multiple reactions were pooled for library preparation (3) to enable SMRT-CCS (more than 20 passes) (4). **b**, Exon–exon junctions identified. **c**, Key for outermost circle of **d** and **e**, representing the sum of changes at each genomic location. **d**, **e**, Concentric circle plots of the *APP* locus depicting IEJs (central lines), deletions (del), insertions (ins) and SNVs sequenced from the brains of five individuals with SAD (**d**) and five

individuals without SAD (**e**). Black dots indicate the abundance of deletions, insertions and SNVs on a $\log_{10}$ scale at the specified location. The Aβ region is highlighted in red and known FAD-associated mutations are circled in red. **f**, **h**, The percentage of unique reads from each exon–exon junction for brains from individuals with SAD (**f**) and without (**h**). **g**, **i**, Number of unique IEJs from each exon–exon combination for brains from individuals with SAD (**g**) and without (**i**). **j**, Eleven FAD-associated mutations were identified in 6 *APP* gencDNAs and *APP*-751. In-frame (red) and out-of-frame (grey) mutations are based on the known *APP* reading frames.

SureSelect targeted DNA pull-down (Extended Data Fig. 2i), which showed unbiased genomic coverage across the entire *APP* genomic locus (all introns and exons including exon 8; Extended Data Fig. 2j). Analysis of DNA from 40,000 neuronal nuclei from individuals with SAD identified all previously detected gencDNA exon–exon junctions excluding exon 8, which is absent from the brain-specific *APP* mRNA splice variants *APP*-751 and *APP*-695 (Fig. 2o; see also Fig. 1e).

Distinct *APP* gencDNAs in SAD

The diversity of gencDNA sequences was assessed by a distinct technical approach, single molecule real-time (SMRT) circular consensus sequencing (CCS), which enables high-certainty, long-read calls to be produced by multiple passes over the same template. gencDNAs were enriched by multiple PCR reactions on small neuronal populations from five individuals with SAD (149 reactions from 96,424 nuclei) and five healthy brains (244 reactions from 162,248 nuclei; Fig. 3a). Samples were pooled for library preparation and SMRT-CCS. Of note, more non-diseased nuclei than diseased nuclei were required to produce sufficient product for sequencing. We identified 6,299 unique sequences (10% in frame; Extended Data Fig. 3a, b) including 45 different IEJs, in neuronal nuclei from the brains of individuals with SAD, and 1,084 unique sequences (12.1% in frame; Extended Data Fig. 3a, c),

including 20 IEJs, in neuronal nuclei from non-diseased brains (Fig. 3b–i).

Critically, both qualitative and quantitative differences in the sequences of gencDNA variants distinguished the brains of individuals with SAD from healthy brains (Fig. 3b–i). Distinctions included gencDNAs with novel IEJs and SNVs (Fig. 3d, e), which were far more prevalent in the brains of individuals with SAD. By contrast, gencD-NAs of the canonical neuronal splice variants, *APP*-751 and *APP*-695, predominated in non-diseased brains, and brains from individuals with SAD showed reduced *APP*-751 and no *APP*-695 (Fig. 3f, h). Notably, 11 SNVs that had been previously published as pathogenic FAD mutations (Fig. 3d, j, Supplementary Tables 3, 5), including the Indiana mutation[12], were present in neurons from individuals with SAD. No FAD mutations were detected in non-diseased brains (Fig. 3e, Supplementary Tables 4, 6).

gencDNA formation in cell lines

APP gencDNAs lacking introns, and the presence of brain-specific iso-forms (*APP*-751, *APP*-695), support origins of gencDNAs from RNAs that must involve reverse transcription. To model gencDNA production in cell lines, we expressed *APP*-751 cDNA (Fig. 4a) in a Chinese hamster ovary (CHO) cell line with endogenous reverse transcriptase
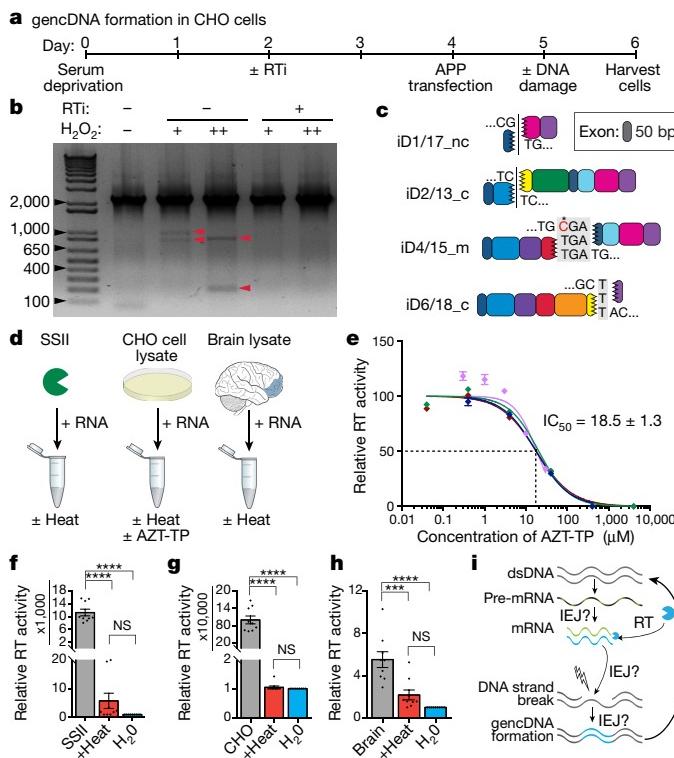
Fig. 4 | Mechanistic studies of gencDNA formation in culture.
a, Timeline of CHO cell experiments. RTi, AZT (100 μM) and ABC (10 μM); transfection, *APP*-751; DNA damage, H_2O_2 at 5 μM (+) and 50 μM (++). **b**, Gel electrophoresis with red arrowheads indicating cloned and sequenced bands. **c**, New induced gencDNA variants. **d**, Reverse transcriptase (RT) activity was analysed in Super Script II (SSII) positive controls, CHO cell lysate, and human brain lysate. **e**, Four independent experiments showed decreased reverse transcriptase activity in CHO cells in response to the RTi azidothymidine triphosphate (AZT-TP). Colours represent individual experiments. **f**–**h**, Relative reverse transcriptase activity in SSII controls, CHO cells, and brain samples (three independent experiments with three biological replicates). Statistical significance was determined using ordinary one-way ANOVA with Sidak's multiple comparisons test. ***$P = 0.003$, ****$P < 0.0001$. NS, not significant (**f**, **g**, $P > 0.9999$; **h**, $P = 0.3095$). Error bars show s.e.m. **i**, Proposed model of reverse transcriptase activity in the formation of gencDNAs.

activity. Initial results did not show gencDNAs, but induction of DNA strand breaks by H_2O_2 produced novel gencDNAs (Fig. 4b, c). Additionally, endogenous reverse transcriptase activity was also required to produce gencDNAs, based on results using the nucleoside reverse transcriptase inhibitors (RTi) abacavir (ABC) and azidothymidine (AZT) (Fig. 4b). Variant RNAs were also dependent on reverse transcriptase activity (Extended Data Fig. 4). Endogenous reverse transcriptase activity was confirmed in CHO cells and further identified in human prefrontal cortex (Fig. 4d–h), consistent with gencDNA production from RNA intermediates and reverse transcription (Fig. 4i).

Increased gencDNAs in SAD and J20 neurons
We further explored the relationships between gencDNAs and SAD using DISH. We examined two gencDNA junctions, DISH$_{16/17}$ and DISH$_{3/16}$, in neurons from six individuals with verified SAD and six non-diseased brains (Fig. 5a–f, Extended Data Fig. 5a–f, Extended Data Table 2; average age 83.5 and 86.7 years, respectively). The number of red foci in neurons from individuals with SAD was three- to fivefold higher than in non-diseased neurons and ranged from 0 to a maximum of 13 in SAD nuclei. Rare foci were observed in non-neuronal nuclei but were not statistically increased in SAD (Fig. 5a–f, Extended Data Fig. 5a–f). Increased gencDNAs in neurons from individuals with SAD raised the question of whether gencDNAs could give rise to toxic

proteins. We therefore tested the cytotoxicity of three *APP* RNA variants (R2/18, R3/14, and R3/16), which were translated in vitro, and found that two of the three variants induced cell death in SH-SY5Y cells (Extended Data Fig. 5g, h).

The J20 mouse model of Alzheimer's disease forms Aβ plaques that accumulate with age. These mice harbour multiple copies of a human *APP* transgene containing the Swedish (K670M/N671L) and Indiana (V717F) mutations, driven by a neuron-specific platelet-derived growth factor-β (PDGF-β) promoter to produce selective, high expression in neurons, with little or no expression in non-neuronal cells[18]. DISH probes for human *APP* did not detect the endogenous mouse locus (Fig. 5g, Extended Data Fig. 6b). DISH$_{3/16}$ identified enriched signals in J20 neuronal nuclei, contrasting with low levels in non-neuronal nuclei from the same mice (Fig. 5g, h, Extended Data Fig. 6a). The more prevalent gencDNA sequence recognized by DISH$_{16/17}$ was also highly enriched in neurons. Notably, DISH$_{16/17}$ demonstrated an age-dependent increase in the area of gencDNA foci over a 2.3-year period, a pattern of change that was not observed in non-neuronal nuclei (Fig. 5i, j, Extended Data Fig. 6c). Use of cells infected with retroviral proviruses containing 0, 1, or 2 copies of the DISH$_{16/17}$ target sequence demonstrated that DISH is semiquantitative and reflects DNA copy numbers (Extended Data Fig. 6d–f). The neuron-selective increase in area of foci occurs during adult life, long after cerebral cortical neurogenesis has ceased[19], further supporting the theory that neuronal gene transcription generates gencDNAs.

Discussion
Human neuronal *APP* gene recombination was identified in brains from healthy controls and individuals with SAD. It was characterized by the mosaic presence of thousands of distinct gencDNA variants that enter neuronal genomic DNA through a process involving *APP* transcription that is influenced by neural activity, DNA strand breaks and reverse transcription (Supplementary Discussion). *APP* gencDNAs bear some resemblance to, but are fundamentally distinct from, processed pseudogenes[20] (non-coding, germline remnants of evolutionarily retrotransposed mRNAs[21,22] that can be active in cancers[23]) and LINE1 repeat elements (which encode an active reverse transcriptase (ORF2)[20] to allow potential retrotransposition in mitotic cells, including within the developing brain[5,9,24–26]). By comparison, *APP* gencDNAs manifest as thousands of distinct genomic variants derived from a cellular gene, contain IEJs and myriad SNVs, can undergo multiple 'retro-insertions' into post-mitotic neuronal genomes, and appear capable of being actively transcribed and translated to produce variant bioactive products that are relevant to both normal and diseased states.

Constitutive mutations or *APP* CNVs are considered causal in FAD and Down syndrome, raising the possibility that previously reported somatic *APP* exonic CNVs contribute mechanistically to SAD[10], which can be explained by the somatic gene recombination identified here. Proof-of-concept data from individuals with SAD identified a marked shift in the forms and abundance of gencDNAs when compared with healthy controls (Figs. 3, 5), including the three- to fivefold increase in gencDNAs in all brains from individuals with SAD examined (Fig. 5). Notably, we identified 11 somatic SNVs that were previously identified as being pathogenic in FAD[12], which were absent from non-diseased controls. Other SNVs, as well as myriad genomic alterations, may also contribute to SAD through both classical and non-classical mechanisms.

Classical mechanisms that support the amyloid hypothesis involve production of toxic Aβ peptides and plaque formation. gencDNAs, including those with FAD-associated mutations, are likely to represent a source of secretase-cleaved substrate for the production of Aβ, as well as a potential source of toxic products that do not require secretase cleavage (Extended Data Fig. 5). Non-ATG translation could potentially occur for out-of-frame variants[27]. Non-classical mechanisms might involve RNA pathologies[28] or maximum limits to gencDNA integration per neuron, beyond which neurodegeneration occurs via genome instability, akin to deleterious mobile elements[29]. The potential diversity of
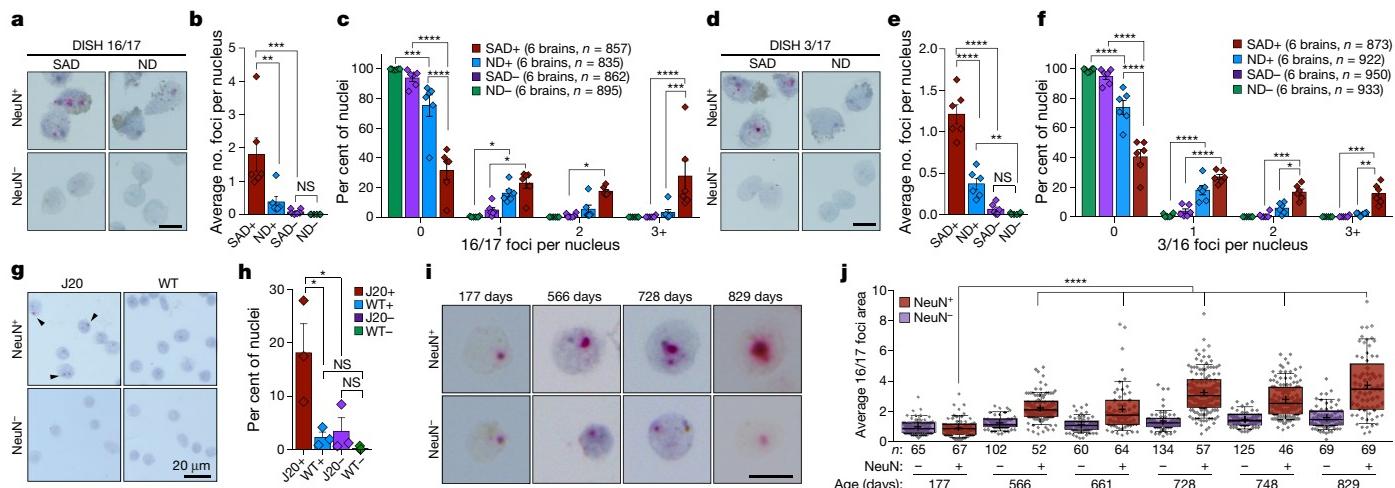
Fig. 5 | Proof-of-concept correlation between gencDNAs and SAD.
a–**f**, Nuclei sorted from the brains of six individuals with SAD and six individuals without SAD were analysed by DISH$_{16/17}$ (**a**–**c**) and DISH$_{3/16}$ (**d**–**f**). **a**, **d**, Representative DISH images. **b**, **e**, Average number of foci per nucleus was increased in neurons from individuals with SAD (one-way ANOVA with Holm–Sidak's multiple comparison test). **c**, **f**, Frequency distributions displaying the percentage of nuclei with 0, 1, 2 and 3 or more (3+) foci (two-way ANOVA with Tukey's multiple comparison test). **g**, Representative DISH$_{3/16}$ of J20 and wild-type (WT) nuclei. **h**, The percentage of nuclei with one or more foci was increased in J20 neurons (ordinary one-way ANOVA with Sidak's multiple comparisons test, J20+

versus WT+ $P = 0.0253$, J20+ versus J20− $P = 0.0371$, WT+ versus WT− $P = 0.9267$, J20− versus WT− $P = 0.9842$). **i**, **j**, Area of DISH$_{16/17}$ foci increased with age in J20 mice. **i**, Representative images of from mice aged 177, 566, 661, 728, 748 and 829 days (one animal each, number of nuclei interrogated is listed below box). **j**, Area of foci shows statistically significant increases with age. +, mean; line, median; box, 75th–25th percentiles; whiskers, 90th–10th percentiles (non-parametric Kruskal–Wallis with Dunn's multiple comparisons test). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. NS, not significant. Detailed P values for **b**, **c**, **e**, **f**, **j** are listed in Extended Data Figs. 5, 6. Error bars show s.e.m. Scale bars, 10 μm unless otherwise noted.

protein variants produced by gencDNAs, and other non-protein mechanisms, may help to explain the failure of therapeutic trials targeting Aβ and related enzymologies, especially those targeting single molecular entities[30].

The largest risk factor for SAD is age, and the age-related increase in gencDNA variants in neurons offers a possible explanation for the decades of life required for SAD to manifest. Neuronal *APP* transcription promotes gencDNA generation both in cell culture and in J20 neurons in vivo (Figs. 4, 5). This is consistent with the increase in *APP* transcription that was previously linked to SAD incidence[31], and the gene encoding the SAD risk factor ApoE4[32]. Notably, the dependence of gencDNA on reverse transcriptase activity might be relevant to the statistical rarity of proven cases of Alzheimer's disease in individuals with HIV infection who are more than 65 years old[33,34], and who have received prolonged, combined anti-retroviral therapy (cART), which includes reverse transcriptase inhibitors. If confirmed, this observation would suggest the immediate use of FDA-approved cARTs or modified combinations containing reverse transcriptase inhibitors to treat SAD, Down syndrome and perhaps FAD. Additionally, processes that produce DNA breaks, such as head injury, that have been linked to Alzheimer's disease[35–37] are consistent with gencDNA production requiring DNA breaks. Thus, gencDNAs and their production have properties relevant to a range of Alzheimer's disease mechanisms and the development of new therapeutic strategies.

The presence of *APP* gencDNAs in non-diseased neurons is likely to reflect the normal roles of *APP*[38], including synaptic function; here, *APP* gencDNAs might provide an increased repertoire of protein species, contributing to synaptic diversity. Additional genes may be transcriptionally modified and genomically retro-inserted in response to selective activities in neuronal populations. Such a mechanism might enable preferential gene re-expression that bypasses splicing or further RNA modification. More broadly, gencDNAs could provide neurons with an activity-dependent mechanism for recording and retaining information over long periods of time, perhaps placing multiple forms of a gene under transcriptional control distinct from a wild-type locus, which could be produced through diverse genomic integration sites that remain to be determined. Such a process could have relevance to known neuronal functions that depend on transcriptional activity,

including Hebbian plasticity[39], synaptic wiring[40], learning and memory[41], and cognition[42]. Thus, gencDNA production may represent both a 'recording' and a 'playback' mechanism for expressing a symphony of variants beyond wild-type gene forms. It would be surprising if *APP* were the only gene to undergo this form of recombination, which might influence distinct, normal brain functions as well as contributing to brain disorders such as Alzheimer's disease.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0718-6.

1. Dreyer, W. J., Gray, W. R. & Hood, L. The genetic, molecular and cellular basis of antibody formation: some facts and a unifying hypothesis. *Cold Spring Harb. Symp. Quant. Biol.* **32**, 353–367 (1967).
2. Hozumi, N. & Tonegawa, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl Acad. Sci. USA* **73**, 3628–3632 (1976).
3. Chun, J. J., Schatz, D. G., Oettinger, M. A., Jaenisch, R. & Baltimore, D. The recombination activating gene-1 (RAG-1) transcript is present in the murine central nervous system. *Cell* **64**, 189–200 (1991).
4. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
5. Rohrback, S., Siddoway, B., Liu, C. S. & Chun, J. Genomic mosaicism in the developing and adult brain. *Dev. Neurobiol.* https://doi.org/10.1002/dneu.22626 (2018).
6. Rehen, S. K. et al. Chromosomal variation in neurons of the developing and adult mammalian nervous system. *Proc. Natl Acad. Sci. USA* **98**, 13361–13366 (2001).
7. Rehen, S. K. et al. Constitutional aneuploidy in the normal human brain. *J. Neurosci.* **25**, 2176–2180 (2005).
8. Westra, J. W. et al. Neuronal DNA content variation (DCV) with regional and individual differences in the human brain. *J. Comp. Neurol.* **518**, 3981–4000 (2010).
9. McConnell, M. J. et al. Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science* **356**, eaal1641 (2017).
10. Bushman, D. M. et al. Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *eLife* **4**, (2015).
11. Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* **8**, 595–608 (2016).

12. Murrell, J., Farlow, M., Ghetti, B. & Benson, M. D. A mutation in the amyloid precursor protein associated with hereditary Alzheimer's disease. *Science* **254**, 97–99 (1991).

13. Hooli, B. V. et al. Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Mol. Psychiatry* **19**, 676–681 (2014).

14. Wiseman, F. K. et al. A genetic cause of Alzheimer disease: mechanistic insights from Down syndrome. *Nat. Rev. Neurosci.* **16**, 564–574 (2015).

15. Rohrback, S. et al. Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing. https://doi.org/10.1073/pnas.1812702115 *Proc. Natl Acad. Sci. USA* (2018).

16. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).

17. Dawkins, E. & Small, D. H. Insights into the physiological function of the β-amyloid precursor protein: beyond Alzheimer's disease. *J. Neurochem.* **129**, 756–769 (2014).

18. Mucke, L. et al. High-level neuronal expression of Aβ$_{1-42}$ in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J. Neurosci.* **20**, 4050–4058 (2000).

19. Ming, G. L. & Song, H. Adult neurogenesis in the mammalian brain: significant answers and significant questions. *Neuron* **70**, 687–702 (2011).

20. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24**, 363–367 (2000).

21. Harrison, P. M., Zheng, D., Zhang, Z., Carriero, N. & Gerstein, M. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res.* **33**, 2374–2383 (2005).

22. Vanin, E. F. Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* **19**, 253–272 (1985).

23. Kalyana-Sundaram, S. et al. Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* **149**, 1622–1634 (2012).

24. Evrony, G. D., Lee, E., Park, P. J. & Walsh, C. A. Resolving rates of mutation in the brain using single-neuron genomics. *eLife* **5**, e12966 (2016).

25. Muotri, A. R. et al. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).

26. Upton, K. R. et al. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **161**, 228–239 (2015).

27. Cleary, J. D. & Ranum, L. P. Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. *Curr. Opin. Genet. Dev.* **26**, 6–15 (2014).

28. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).

29. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–801 (1984).

30. Egan, M. F. et al. Randomized trial of verubecestat for mild-to-moderate Alzheimer's disease. *N. Engl. J. Med.* **378**, 1691–1703 (2018).

31. Brouwers, N. et al. Genetic risk and transcriptional variability of amyloid precursor protein in Alzheimer's disease. *Brain* **129**, 2984–2991 (2006).

32. Huang, Y. A., Zhou, B., Wernig, M. & Sudhof, T. C. ApoE2, ApoE3, and ApoE4 differentially stimulate APP transcription and Aβ secretion. *Cell* **168**, 427–441. e21 (2017).

33. Turner, R. S. et al. An individual with human immunodeficiency virus, dementia, and central nervous system amyloid deposition. *Alzheimers Dement. (Amst.)* **4**, 1–5 (2016).

34. Centers for Disease Control and Prevention. *HIV Surveillance Report, 2016* Vol. 28 http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html (2017).

35. Suberbielle, E. et al. Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-β. *Nat. Neurosci.* **16**, 613–621 (2013).

36. Mortimer, J. A., French, L. R., Hutton, J. T. & Schuman, L. M. Head injury as a risk factor for Alzheimer's disease. *Neurology* **35**, 264–267 (1985).

37. Pravdenkova, S. V., Basnakian, A. G., James, S. J. & Andersen, B. J. DNA fragmentation and nuclear endonuclease activity in rat brain after severe closed head injury. *Brain Res.* **729**, 151–155 (1996).

38. Nhan, H. S., Chiang, K. & Koo, E. H. The multifaceted nature of amyloid precursor protein and its proteolytic fragments: friends and foes. *Acta Neuropathol.* **129**, 1–19 (2015).

39. Guzman-Karlsson, M. C., Meadows, J. P., Gavin, C. F., Hablitz, J. J. & Sweatt, J. D. Transcriptional and epigenetic regulation of Hebbian and non-Hebbian plasticity. *Neuropharmacology* **80**, 3–17 (2014).

40. Hattori, D., Millard, S. S., Wojtowicz, W. M. & Zipursky, S. L. Dscam-mediated cell recognition regulates neural circuit formation. *Annu. Rev. Cell Dev. Biol.* **24**, 597–620 (2008).

41. Madabhushi, R. et al. Activity-induced DNA breaks govern the expression of neuronal early-response genes. *Cell* **161**, 1592–1605 (2015).

42. West, A. E. & Greenberg, M. E. Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb. Perspect. Biol.* **3**, a005744 (2011).

Author contributions J.C. conceived the project. M.-H.L. and J.C. designed and analysed experiments. M.-H.L. (nuclei sorting, RT–PCR, Southern blot, cloning and Sanger sequencing, genomic DNA PCR, RNA and DNA in situ hybridization, targeted DNA pull-down and Illumina sequencing, SMRT sequencing, gencDNA induction in culture and cytotoxicity assay), B.S. (SMRT sequencing), R.R. (genotyping, breeding and maintenance of mouse lines and DNA concatemer preparation and virus infection), W.J.R. (nucleus sorting, in vitro reverse transcriptase activity assay and western blot), C.P. (genomic DNA PCR) and G.K. (RNA and DNA in situ hybridization) completed experiments. M.-H.L. and G.E.K. performed statistical analyses and made figures. B.S., I.S., C.S.L. and T.L. performed informatics analyses. M.-H.L., B.S., G.E.K. and J.C. prepared the manuscript. Key experiments were repeated by G.E.K., W.J.R., C.P. and other researchers in our laboratory.

Competing interests Sanford Burnham Prebys Medical Discovery Institute has filed the following patent applications on the subject matter of this publication: (1) PCT application number PCT/US2018/030520 entitled "Methods of diagnosing and treating Alzheimer's disease" filed 1 May 2018, which claims priority to US provisional application 62/500,270 filed 2 May 2017; and (2) US provisional application number 62/687,428 entitled "Anti-retroviral therapies and reverse transcriptase inhibitors for treatment of Alzheimer's disease" filed 20 June 2018.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0718-6.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0718-6.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to J.C.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Human brain tissue and J20 mice. Fresh frozen human brain tissue was provided by the University of California San Diego (UCSD) Alzheimer's Disease Research Center (ADRC) and the University of California Irvine (UCI) Institute for Mind Impairments and Neurological Disorders (MIND).

J20 transgenic mice (B6/Cg-Tg(PDGFB-APPSwInd)20Lms/2JMmjax) were purchased from The Jackson Laboratory and housed in IACUC approved animal facilities in accordance with applicable laws and regulations at Sanford Burnham Prebys Medical Discovery Institute. Sex (F/M) and age (days) of mice used for experiments are listed: F177, M566, M661, M661, M728, F748, F829, and M861. Sample sizes were estimated based upon preliminary data without additional statistics. Samples were allocated randomly and in situ hybridization quantification was blinded for statistical assessments.

Nucleus extraction and FANS. Human and mouse brain nuclei were isolated as described previously[10]. For in situ hybridization analyses, isolated nuclei were fixed in 1:10 diluted buffered formalin (Fisher Healthcare) for 5 min. Prior to sorting, fixed or unfixed nuclei were then labelled with anti-NeuN rabbit monoclonal antibody (1:800) (Millipore, Germany) and Alexa Fluor 488 donkey anti-rabbit IgG (1:500) (Life Technologies, Carlsbad, CA), and counterstained with propidium iodide (PI; 50 µg/ml) (Sigma, St. Louis, MO). For DNA analyses, RNase A (100 µg/ml) was included with all subsequent steps after initial nuclei isolation, including primary and secondary antibody incubations. Diploid NeuN-positive and negative nuclei were gated by PI and immunofluorescence, and sorted into appropriate populations for RT–PCR, genomic DNA PCR, or in situ hybridization. FANS was performed on a FACSAria Fusion (BD Biosciences, Franklin Lakes, NJ) or with a FACS-Aria II (BD Biosciences, Franklin Lakes, NJ).

RNA extraction and RT–PCR. All RNA extractions from 50-nuclei populations (NeuN-negative and positive) and bulk tissues were performed using Quick-RNA MicroPrep (Zymo Research, Irvine, CA) and RNAeasy Mini kits (Qiagen, Valencia, CA) according to the manufacturer's protocols. OneStep Ahead RT–PCR (Qiagen, Valencia, CA) was used for RT–PCR with *APP* 1–18 primer sets (Supplementary Table 1) according to the manufacturer's protocol. Oligo-(dT)$_{20}$ primer was used to prime the cDNA library as indicated. Low annealing stringency PCR was carried out with the following thermal cycling steps for 40 cycles: 95 °C 15 s, 55 °C for *APP* and 52 °C for *PSEN1* 15 s, and 68 °C 2.5 min for *APP* and 2 min for *PSEN1*.

Southern blotting. RT–PCR products were run on agarose gel, denatured, and transferred to a positively charged nylon membrane. UV-crosslinked membranes were incubated with denatured and purified ^{32}P-labelled *APP* cDNA probes at 42 °C overnight. Blots were washed four times with increasing washing stringency and temperature according to established protocols. Images were developed on a Typhoon (GE Healthcare Life Sciences) or Fujifilm FLA-5100 phosphorimager.

DNA extraction and genomic DNA PCR. DNA extraction from isolated neuronal nuclei populations was performed via isopropanol precipitation. In brief, nuclei were incubated with proteinase K in 550 µl PK buffer (50 mM Tris pH 8.0, 0.1 M EDTA, 0.1 M NaCl, 1% SDS) overnight at 55 °C. Samples were then treated with RNase cocktail enzyme mix (ThermoFisher, Waltham, MA) for 2 h, followed by addition of 250 µl saturated NaCl. After centrifugation, supernatant was used for DNA precipitation by isopropanol and washed three times with 70% ethanol. DNAeasy and QIAamp DNA Mini kits (Qiagen, Valencia, CA) were also used according to the manufacturer's instructions. Purified DNA was stored at −20 °C for future use. High annealing stringency PCR was performed using either the FastStart PCR master mix (Sigma, St. Louis, MO) with PCR cycle settings: 95 °C 30 s, 65 °C for 30 s, and 72 °C 2.5 min for 40 cycles, or the Platinum SuperFi DNA polymerase (ThermoFisher, Waltham, MA) with cycle settings: 98 °C 10 s, 65 °C for *APP* and 52 °C for *PSEN1* for 10 s, and 72 °C 1.5 min for *APP* and 1 min for *PSEN1* for 40 cycles. Primer sequences are listed in Supplementary Table 1.

DNA in situ hybridization and RNA in situ hybridization. For DISH pretreatment, sorted nuclei were dried on Plus Gold slides (Fisher Scientific, Pittsburgh, PA). Nuclei were then treated with RNase cocktail enzyme mix (RNase A + RNase T1, 1:50) (ThermoFisher, Waltham, MA) at 40 °C for 60 min, followed by fixation in 1:10 dilution buffered formalin at room temperature for 5 min. After being washed with distilled water twice, slides were treated with hydrogen peroxide at room temperature for 10 min, target retrieval reagent at 95 °C for 15 min, followed by protease treatment at 40 °C for 10 min. Restriction enzyme was applied after protease treatment for 2 h as necessary in negative control experiments. DNA was then denatured (2×SSC, 70% formamide and 0.1% sodium dodecyl sulfate) at 80 °C for 20 min. After cooling the slides to room temperature, BaseScope probes were applied and incubated with nuclei at 40 °C overnight. Samples were then prepared for signal development. For RISH pretreatment, 10 µm fresh frozen human tissue sections were fixed in 1:10 dilution buffered formalin on ice for 10 min. After washing with PBS twice, tissue sections were placed in serial diluted ethanol 50%, 70% and 100%, 5 min for each step. Slides were then treated with hydrogen peroxide at room temperature for 10 min, followed by protease treatment at room temperature for 20 min. BaseScope probes were then incubated with tissue sections

at 40 °C for 2 h. Hydrogen peroxide, 10× target retrieval buffer, proteases, custom BaseScope probes (Supplementary Table 1), and BaseScope reagent kit-RED used for signal development were all purchased from Advanced Cell Diagnosis (ACD, Newark, CA). Duplex BaseScope reagent kit was also purchased from ACD. Nuclei and tissue sections were counterstained with haematoxylin. Zeiss AX10 Imager, M2 microscope and ZEN2 software were used for image acquisition. Images were thresholded, and foci number or size were quantified using ImageJ for statistical analysis.

Agilent SureSelect hybridization enrichment and sequencing. The method is graphically represented in Extended Data Fig. 2i. Nuclei were isolated from human frontal cortex, labelled for NeuN, and NeuN-positive nuclei were isolated via FANS. Genomic DNA was extracted and fragmented into ~1.2 kb using sonication (Covaris, Woburn, MA). End repair reactions were performed and Illumina sequencing adaptors were ligated to genomic DNA. Library-prepped DNA was hybridized with custom Agilent SureSelect probes designed against the entire *APP* locus, including introns. Purified *APP*-containing genomic DNA sequences were then sequenced on an Illumina NextSeq (Illumina, San Diego, CA). Sequences were aligned to the human reference genome (GRCh38) using STAR (version 2.5.3a) with the settings: --outSAMattributes All --outFilterScoreMinOverLread 0.8 --outSJfilterCountTotalMin 1 1 1 1. Duplicate reads were marked and removed using Picard (version 2.1.1). Reads were then informatically analysed using IGV, the UCSC Genome Browser, and a custom imaging pipeline built in R.

SMRT sequencing. Neuronal genomic DNA was isolated as described above and used for *APP* PCR and nested *APP* PCR. Platinum SuperFi DNA polymerase with 100× higher fidelity compared to native *Taq* (Invitrogen, Platinum SuperFi DNA Polymerase) was used under high annealing stringency (98 °C 10 s, 65 °C 10 s, and 72 °C 1.5 min, for 30 cycles). An aliquot of the first PCR product was used as a DNA template for nested PCR reactions. Multiple PCR reactions were pooled (149 reactions for Alzheimer's disease and 244 reactions for non-diseased) and purified by DNA Clean and Concentrator-5 (Zymo Research, Irvine, CA) for SMRT sequencing library preparation. PCR amplicons were repaired using SMRTbell template prep kit version 2.0 (PacBio) and purified using AMPure PB beads (PacBio, Menlo Park, CA). Adapters were ligated to DNA to create SMRTbell libraries. Sequencing polymerase was annealed and the SMRTbell library was loaded using Magbead binding. Raw bam sequencing files were converted to fastq format using the CCS algorithm in the SMRTLink software tool kit from PacBio. In CCS, reads were included in the fastq file only if 1) there were more than 20 passes of the sequencing polymerase over the DNA insert in the zero mode waveguide well and 2) the predicted accuracy in SMRTLink was calculated to be greater than 0.9999. These cutoffs generated ultra-high accuracy reads, the median Phred score of reads used was 93[43–46], representing 99.999999% accuracy, with further quality filtering steps applied in our informatic analysis. This SMRT sequencing is comparable in fidelity to Sanger sequencing[43,45].

Genomic data analyses with customized bioinformatic pipelines. Novel algorithms were developed to detect and analyse exon rearrangement in genes of interest. The algorithms were specifically designed to analyse long-read sequences generated by the Pacific Biosciences Sequel platform. A series of quality control (QC) procedures were performed before sequence processing to ensure high quality of the reads being analysed.

Quality control: consensus sequence and read quality. PacBio circular consensus sequence (CCS) reads with fewer than 20 passes were filtered out to ensure overall sequence quality. Quality score distributions were examined: for *APP* gene PCR enriched sequences, average median read-wide Phred score was 93. In gencDNA analyses, we included only reads that met a mean Phred cutoff of >85.

Quality control: sequencing artefacts. Owing to the intrinsic limits of PacBio SMRT sequencing technology, errors in homopolymers (that is, sequence ATTTG could be read as ATTTTG or ATTG in addition to ATTTG) are specially handled with a method that combines quality score information and reference sequence at the beginning of the homopolymer. The FASTQ files encoded uncertainty in the homopolymer run length in the first Phred score of each run. If this Phred score was lower than our threshold of 30, then this position was marked as a likely sequencing artefact and not a real variant.

PCR primer filter. The reads were checked to ensure the correct start and end sites with forward and reverse PCR primer sequences. BLAST (command line tool 'blastn' 2.6.0+) was used to align primer sequences in either orientation to each read with word size 13, gap open penalty 0, and gap extension penalty 2. Any read in which both primers were not detected was filtered out. Furthermore, reads on the negative strand were reverse complemented in this step. BLAST seed length was optimized to avoid ambiguity and ensure sensitivity.

Alignment to APP reference sequences. The Ensembl reference sequence for *APP* was downloaded from the GRCh38 reference human genome assembly using the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) with RefSeq accession number NM_000484.3. Because the PCR primers begin at the start codon and end with the stop codon, sequences of exons 1 and 18 were trimmed to these

positions so that only the coding sequence of each of the 18 exons was kept and stored as a FASTA file. Then, we used BLAST to look for local alignment between 18 exons and each quality-filtered CCS read; blastn parameters used: –outfmt 6, -wordsize 25, -gapopen 0, -gapextend 2. We used the resulting alignment coordinates to mark regions of each read covered by exons. This allowed us to analyse exon arrangements, lengths and patterns of exon–exon junctions.

SNV and INDEL analysis. First, we used reference sequences of *APP* exons to replace low quality individual nucleotides (potential homopolymer runs and other errors) within each read with their reference *APP* exon counterpart. Then, we analysed BLAST local alignments between each exon (or part of an exon) and the read sequence, nucleotide by nucleotide, to look for alignment mismatches. If the mismatch position was a different nucleotide, we assigned it as a single nucleotide variant (SNV); if the mismatch position was a hyphen in the exon sequence, we assigned it as an insertion and if the mismatch position was a hyphen in the read sequence we assigned it as a deletion.

gencDNA production in culture. The method is graphically represented in Fig. 4a. CHO cells were serum-deprived for 2 days, followed by addition of reverse transcriptase inhibitors AZT (100 μM) and ABC (10 μM) (Tocris, Minneapolis, MN) until the end of the experiment. The medium was changed daily with fresh reverse transcriptase inhibitors. Cells were transfected with *APP*-751 driven by the *CAG* promoter by GenJet (SignaGen Laboratories, Gaithersburg, MD) on day 3, then on day 4, cells were treated with 0 μM, 5 μM, or 50 μM hydrogen peroxide (Fisher Scientific) for 2 h. After 1 day, cells were collected and genomic DNA was extracted for PCR analysis.

In vitro reverse transcriptase activity assay. Lysates were prepared in reverse transcriptase disruption buffer[47], and contained cOmplete, EDTA-free protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO) and PhosSTOP phosphatase inhibitors (Sigma-Aldrich, St. Louis, MO). The assay was performed essentially as described[48], except the assay was separated into two parts. One microgram of extract was used in the reverse transcription step of the assay. In addition, Primer A was used in the reverse transcription reaction cocktail instead of Primer B (Supplementary Information1). Reverse transcription was carried out at 37 °C for 45 min, followed by 15 min at 70 °C.

The reverse transcriptase product of this first step was assayed in triplicate by quantitative PCR. Levels of reverse transcription activity were determined by the Delta Cq method, compared to the activity in negative controls (water and no nucleotides), which were given Cq scores of 40. 100,000 picounits of SuperScript II Reverse Transcriptase (ThermoFisher Scientific) were used as a positive control for the assay.

Lysates for heat inactivation experiments were incubated for 15 min at 70 °C before the reverse transcription step. For inhibitor experiments, lysates were incubated with inhibitor in the presence of all the components of the reaction except for dNTPs. After 10 min at room temperature, dNTPs were added and the reaction was incubated at 37 °C as above. AZT-TP was purchased from TriLink Biotechnologies (San Diego, CA).

Construction and retroviral transduction of synthetic human *APP* sequence targets. Phosphorylated oligonucleotides (Integrated DNA Technologies) composed of human *APP* target sequences with BamHI and BglII restriction sites on the 5′ ends were annealed and ligated into the BamHI site of the retroviral expression vector S-003-AB LZRSpBMN-linker-IRES-EGFP. All primer sequences for construction are listed in Supplementary Table 1. Single and concatamerized oligonucleotide inserts were identified by PCR using primers flanking the BamHI insertion site, and identified clones were sequenced to confirm insert copy number (Genewiz, La Jolla, CA). Helper-free ecotropic virus was produced by transfecting DNA constructs (Lipofectamine 2000, Thermo Fisher Scientific) with single or multiple copies of the oligonucleotide inserts into the retrovirus packaging line Phoenix-ECO. Forty-eight hours after transfection, retroviral supernatants were harvested, and 2 ml of selected virus was used for transduction of NIH-3T3 cells in 6-well plates. Retroviral transduction was carried out by removing the cell growth medium, replacing it with 2 ml retroviral supernatant containing 4 μg/ml

polybrene, and spinning at 25 °C for 1 h at 2,800 r.p.m. Forty-eight hours after transduction, the percentage of GFP+ cells, as identified by flow cytometry, was used to evaluate transduction efficiency.

Cell culture. NIH-3T3 (DMEM, 5% FBS), CHO-K1 (RPMI, 10% FBS), SH-SY5Y (DMEM/F12, 10% FBS), IMR-90 (DMEM, 10% FBS) and HEK-293 (DMEM, 10% FBS) cells were purchased from ATCC and maintained at 37 °C under 5% CO_2. Although HEK-293 is in the database of commonly misidentified cell lines, this cell line was used to show protein expression of our targets in mammalian cells and as gencDNA negative controls. For the indicated purpose, there is no concern regarding this cell line in the manuscript. Cells from ATCC and all reagents used were verified to be mycoplasma free.

Western blot. Cells were harvested in RIPA buffer (100 mM Tris–HCl, pH 7.6, 250 mM NaCl, 0.1% sodium dodecyl sulfate, 0.2% deoxycholic acid, 0.5 mM dithiothreitol, 1 mM EDTA, 0.5% NP-40 and 1% Triton X-100), and proteins in each lysate were analysed using rat monoclonal anti-HA antibody (Clone 3F10, Roche) and horseradish peroxidase-conjugated goat anti-rat secondary antibody (Cell Signaling). Enhanced chemiluminescent substrate (Millipore) for the reaction was added, and the signal was detected by BioRad bioimaging system.

Variant toxicity assay. SH-SY5Y cells were transfected with *APP* RNA variants by lipofectamine LTX (Life Technologies) overnight and further cultured under serum-deprived conditions for 7 days. Cell viability was determined by WST1 reagent (Roche Applied Science) according to the manufacturer's protocol. In brief, cells in 96-well plates were incubated with 100 μl WST1 reagent and culture medium in a ratio of 1:10 (v/v) per well at 37 °C for 2 h. The absorbance at 440 nm of samples normalized by a background control was measured by microplate ELISA reader.

Statistics and reproducibility. All statistical analyses were completed using Prism Version 7 (GraphPad). Specific tests, number of data points (*n*), and *P* values are reported in Figures, figure legends or Extended Data Figures. All experiments in Figures and Extended Data Figures were repeated at least three times (independent experiments) unless specified otherwise in the figure legends.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Fastq files of SMRT sequences performed on a PacBio Sequel and Illumina sequences on NextSeq500 have been deposited in NCBI Sequence Read Archive (BioProject ID: PRJNA493258). The PacBio produced RNA-seq data sets from whole brain and temporal lobe supporting the findings of this study are available at https://www.pacb.com/blog/data-release-alzheimer-brain-isoform-sequencing-iso-seq-dataset, and from the authors upon reasonable request and with permission of PacBio, respectively. The source codes of the customized algorithms are available on GitHub (https://github.com/christine-liu/exonjunction and https://github.com/taolonglab/varccs).

43. Hebert, P. D. N. et al. A Sequel to Sanger: amplicon sequencing that scales. *BMC Genomics* **19**, 219 (2018).
44. Eid, J. et al. Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
45. Zhang, H., Susanto, T. T., Wan, Y. & Chen, S. L. Comprehensive mutagenesis of the fimS promoter regulatory switch reveals novel regulation of type 1 pili in uropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **113**, 4182–4187 (2016).
46. Roberts, R. J., Carneiro, M. O. & Schatz, M. C. The advantages of SMRT sequencing. *Genome Biol.* **14**, 405 (2013).
47. Lovatt, A. et al. High throughput detection of retrovirus-associated reverse transcriptase using an improved fluorescent product enhanced reverse transcriptase assay and its comparison to conventional detection methods. *J. Virol. Methods* **82**, 185–200 (1999).
48. Ma, Y. K. & Khan, A. S. Evaluation of different RT enzyme standards for quantitation of retroviruses using the single-tube fluorescent product-enhanced reverse transcriptase assay. *J. Virol. Methods* **157**, 133–140 (2009).
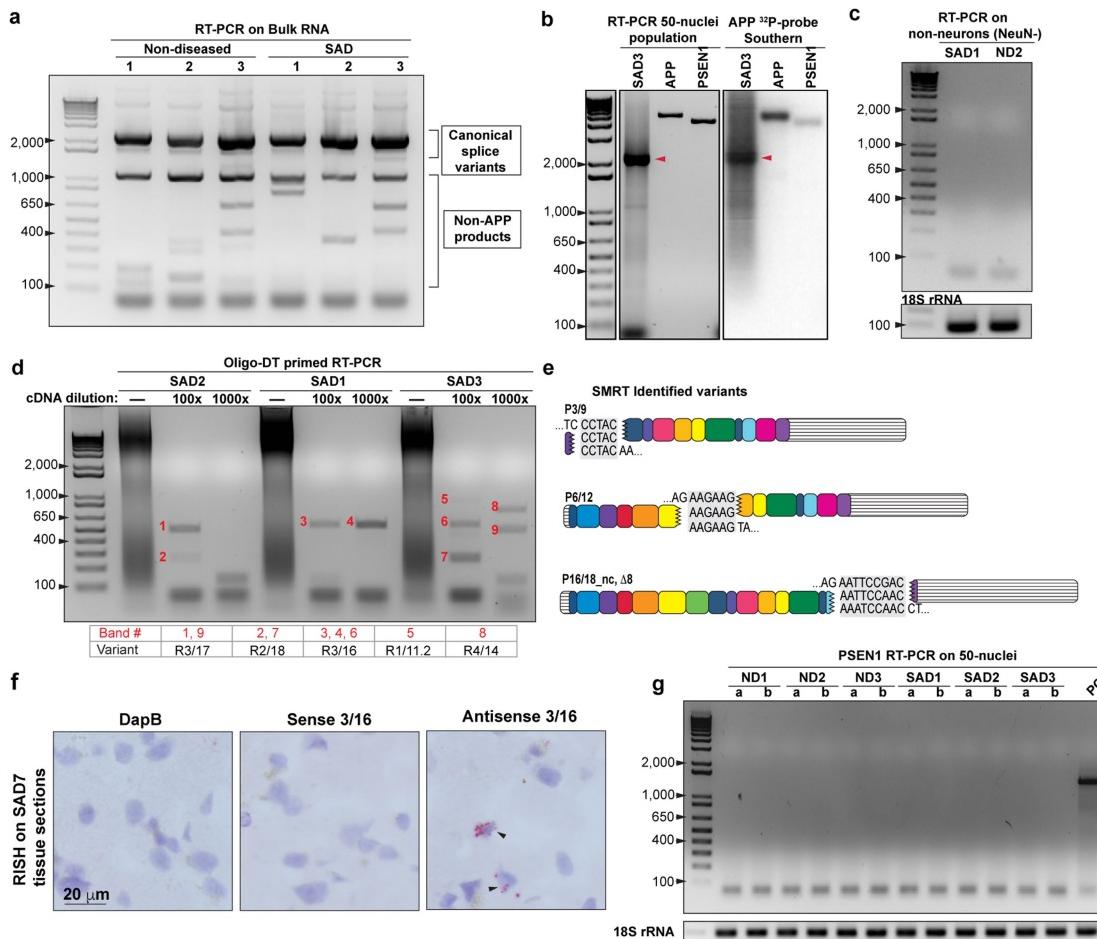
Extended Data Fig. 1 | RT–PCR on bulk and sorted nuclei, and RISH. a, RT–PCR products from bulk brain tissue samples from three individuals with SAD and three without. Canonical *APP* splice variants and non-*APP* products were identified. **b**, Representative gels showing the presence of canonical *APP* splice variants (red arrows, $n = 2$ independent experiments). **c**, No *APP* variants were identified in NeuN-negative nuclei from individuals with or without SAD. The *18S* rRNA control verified the presence of RNA. Novel *APP* RNA variants were identified from oligo-dT primed cDNA libraries from 50-cell populations of neuronal nuclei

(d, $n = 3$ biological replicates) and brains from individuals with Alzheimer's disease (**e**, commercially produced PacBio cDNA libraries). **f**, RISH$_{3/16}$ signal from antisense probes showed cytoplasmic distribution of *APP* 3/16 RNA. Negative control sense probes and a probe targeting the bacterial gene *DapB* showed no signal. **g**, *PSEN1* RT–PCR on populations of 50 nuclei from the brains of three individuals with SAD and three without showed no *PSEN1* RNA variants. The positive control (PC) is amplified from RNA extracted from bulk brain tissue. 18S rRNA control verified the presence of RNA.
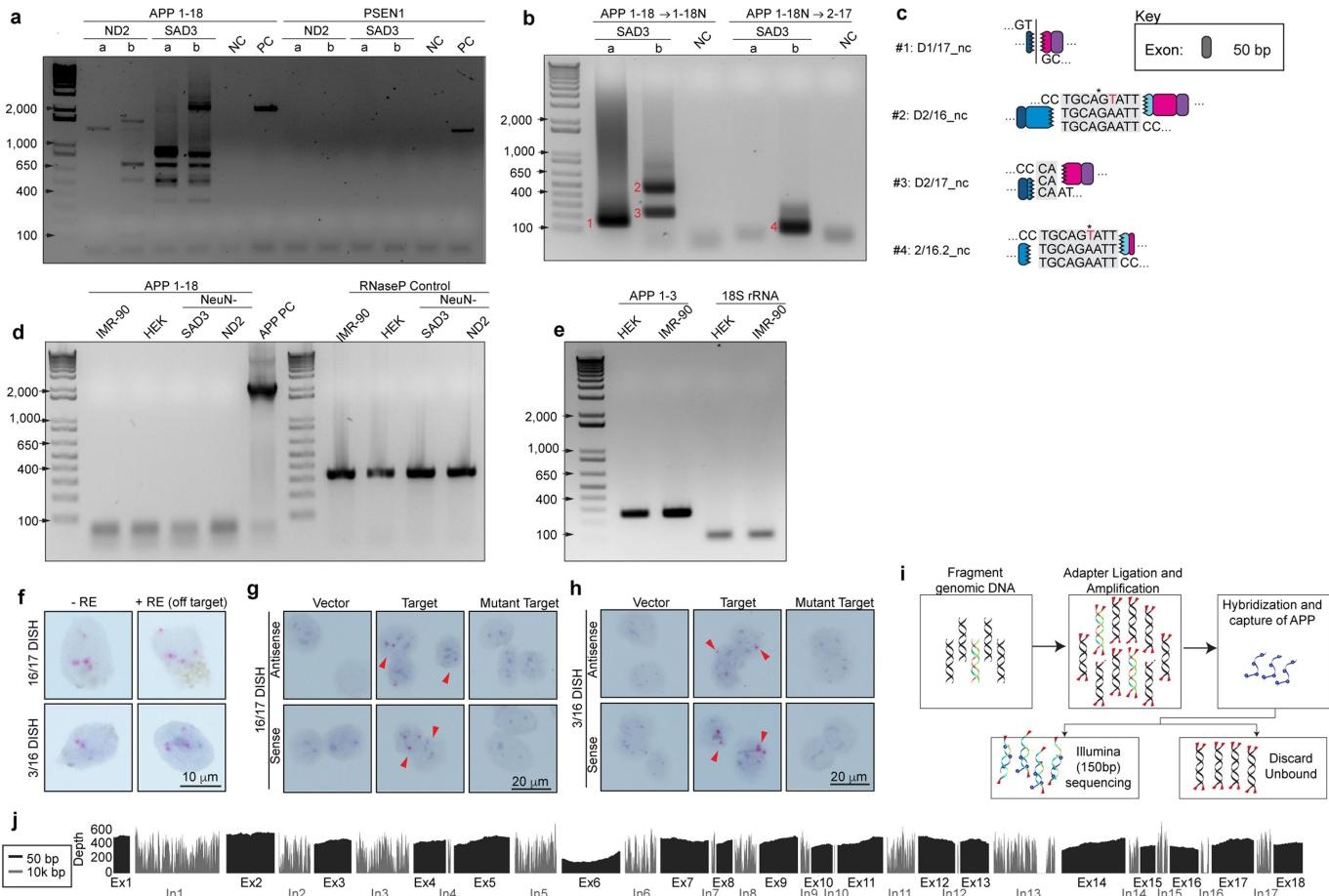
Extended Data Fig. 2 | *APP* gencDNA detection by genomic DNA PCR, DISH, and targeted genomic pull-down. a, Duplicate gel from Fig. 2b, with more sensitive thresholds to show the clear absence of *PSEN1* bands. **b**, Nested PCR was used with alternative *APP* primers (three total sets: *APP* 1–18, *APP* 1–18N, and *APP* 2–17). **c**, Cloning and Sanger sequencing of indicated bands (red numbers in **b**) revealed novel *APP* gencDNAs (see Fig. 1 for legend and nomenclature). **d**, *APP* 1–18 DNA PCR showed no products in non-neuronal cell types: IMR-90 (human lung fibroblast), HEK (human embryonic kidney) and non-neuronal (NeuN-negative) genomic brain DNA from individuals with and without SAD. RNaseP was used as a positive control. **e**, *APP* mRNA is expressed in HEK-293

and IMR-90 cells; *18S* rRNA used as a positive control. **f**, Digestion with the off-target restriction enzyme XbaI did not affect DISH$_{3/16}$ or DISH$_{16/17}$ signals. **g, h**, Synthetic DNA containing 16/17 (**g**) or 3/16 (**h**) target sequences (target), or wild-type human genomic *APP* sequences lacking IEJs and exon–exon junctions (mutant target) were introduced by retroviral transduction into NIH-3T3 cells. DISH$_{16/17}$ and DISH$_{3/16}$ signals from both sense and antisense probes were detected only in target infected cells. **i**, Schematic of Agilent SureSelect targeted DNA pull-down. **j**, Agilent SureSelect hybridization enrichment targeted the entire genomic locus of *APP* and showed unbiased sequencing depth across the full genomic locus. Exons and introns are shown on two scales.
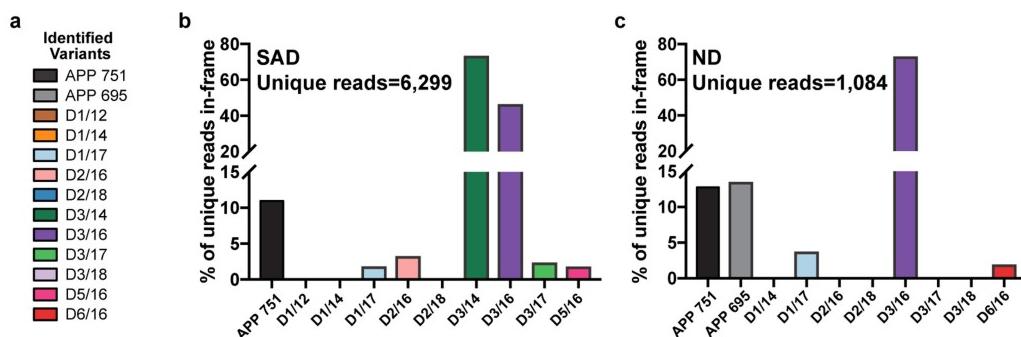
Extended Data Fig. 3 | *APP* gencDNA reading frame analysis. a, Colour key for all gencDNAs with junctions identified by SMRT sequencing. **b**, **c**, Percentage of unique in-frame reads from brains of individual with SAD (**b**; 6,299 unique reads) or without SAD (**c**; 1,084 unique reads).
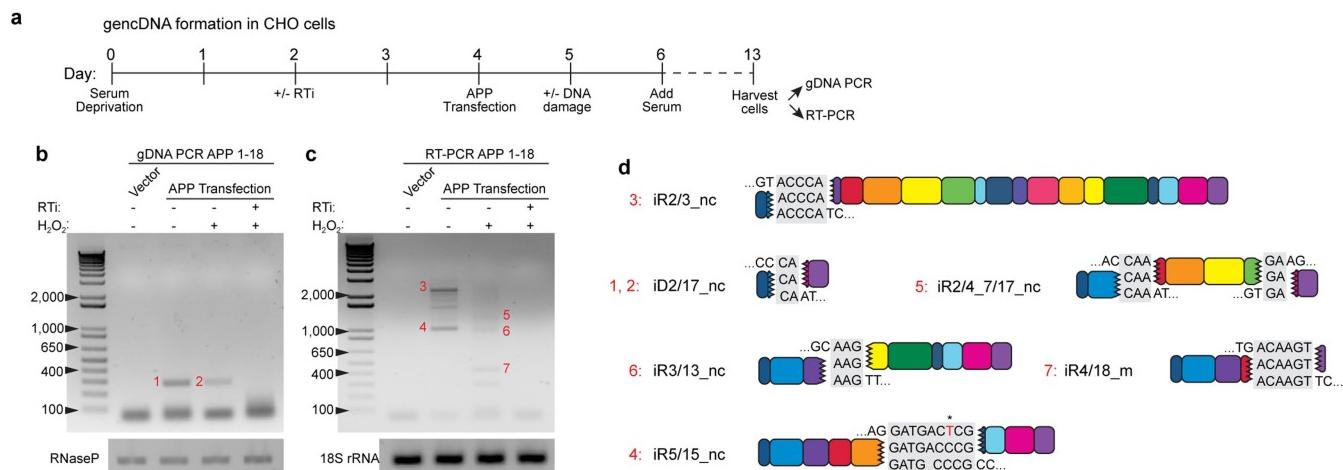
Extended Data Fig. 4 | *APP* gencDNA and RNA variant formation in CHO cells. a, Time line of CHO cell experiments modified from Fig. 4a. After transfection and gencDNA induction, serum was added and CHO cell cultures were passaged for 7 days. Cells were harvested, and DNA and RNA were extracted for analyses. **b**, **c**, PCR of genomic DNA (**b**; gDNA) and RT–PCR with APP 1 and 18 primers (**c**; $n = 2$ independent experiments). Note that APP plasmid is no longer detected (compare to Fig. 4b). DNA breaks during cell proliferation might contribute to variant formation in cells without DNA damage (no H_2O_2). Reverse transcriptase inhibitor (RTi, AZT + ABC) treatment prevents formation of *APP* RNA variants, indicating the dependence of RNA variants on gencDNAs. **d**, Induced *APP* variants with IEJs observed in **b**, **c**.
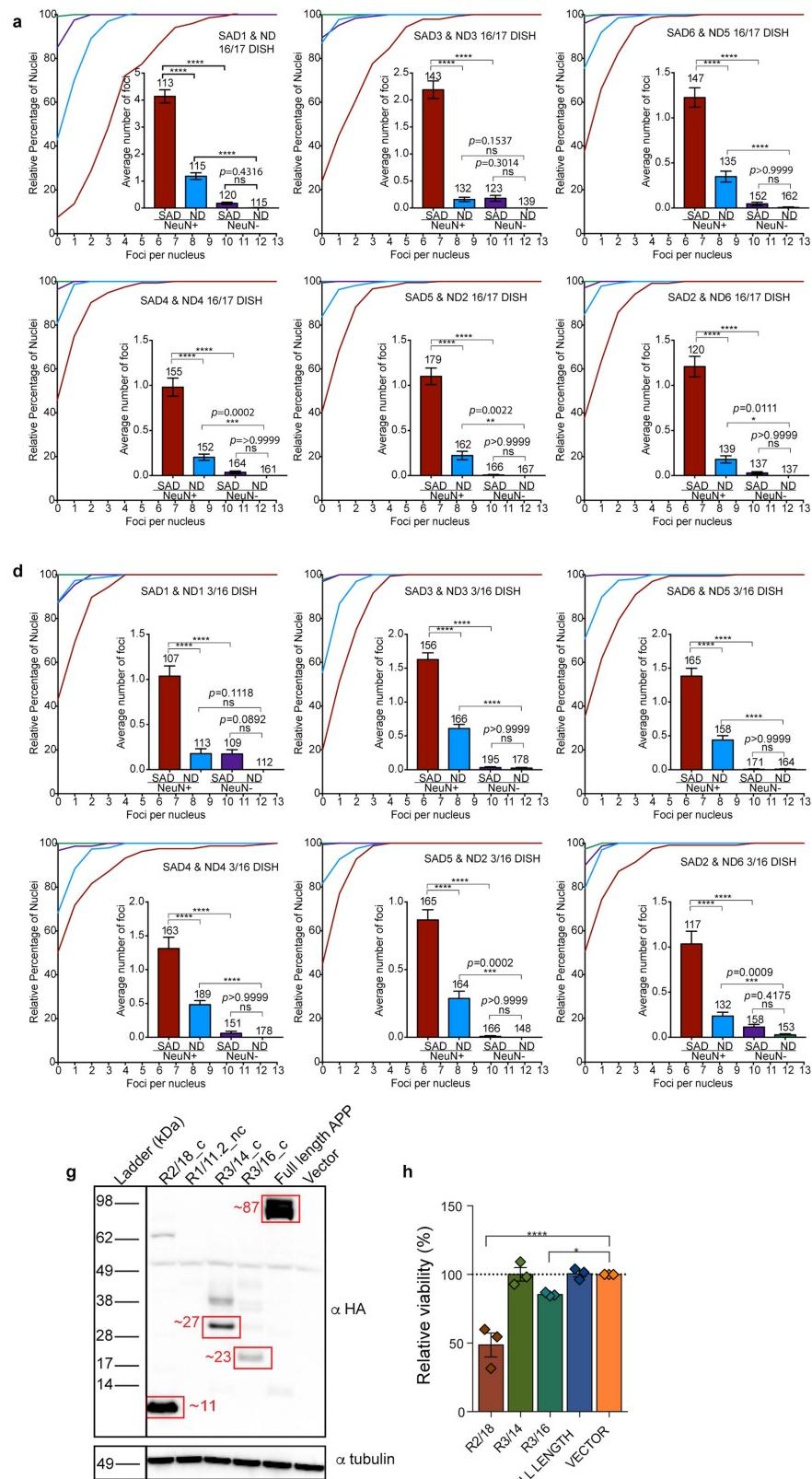
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Data from six individual brains for each brain from individuals with or without SAD represented as averages in Fig. 5, and variant cytotoxicity. a, **d**, Nuclei sorted from cortices of six individuals with SAD and six without were analysed by $DISH_{16/17}$ (**a**) and $DISH_{3/16}$ (**d**). Cumulative frequency distribution plots and average numbers of foci per nucleus show statistical significance (non-parametric Kruskal–Wallis test with Dunn's correction for multiple comparisons) between all paired brain sets. Numbers above bars indicate number of nuclei analysed. NS, not significant. Error bars show s.e.m. **b**, **c**, **e**, **f**, Detailed P values for Fig. 5b (**b**), Fig. 5c (**c**), Fig. 5e (**e**) and Fig. 5f (**f**).

g, *APP*-751, three coding and one non-coding *APP* variant in constructs containing haemagglutinin (HA) tags were transfected into HEK-293 cells. Cell lysates from all three coding variants and full-length *APP*-751 displayed protein products of the expected size by western blot. α-tubulin was used as a loading control. **h**, Three coding *APP* variants were transfected into SH-SY5Y cells individually and cell viability was measured by WST-1 seven days after transfection under serum-deprived conditions. Means of three independent experiments were analysed using ordinary one-way ANOVA with uncorrected Fisher's LSD for multiple comparisons (*$P = 0.0477$, ****$P < 0.0001$).
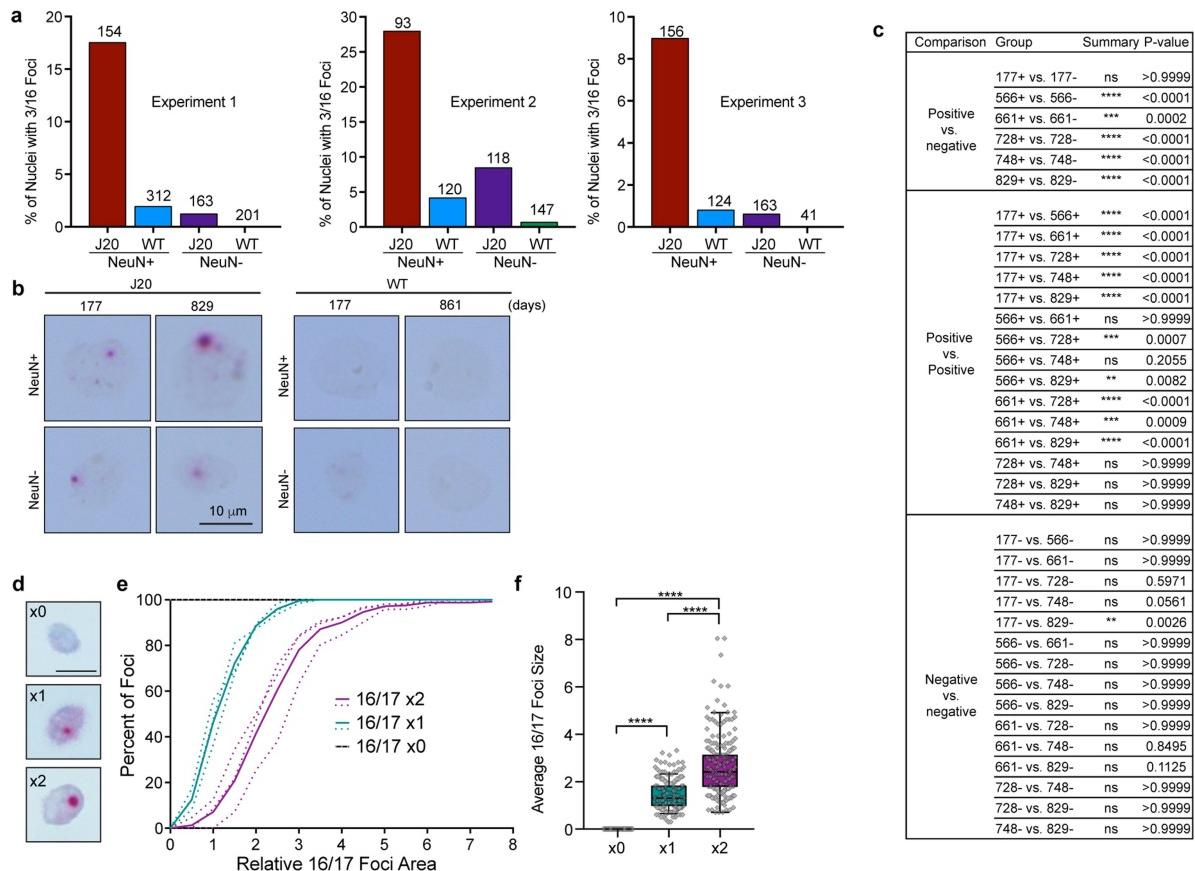
Extended Data Fig. 6 | DISH$_{3/16}$ and DISH$_{16/17}$ data analyses. a, DISH$_{3/16}$ data from individual J20 and wild-type mouse cortices represented as an average in Fig. 5h; numbers above bars represent number of nuclei analysed. **b,** No DISH$_{16/17}$ signal was detected in wild-type mouse nuclei. **c,** Detailed statistical significance of DISH$_{16/17}$ signal across all mice in Fig. 5j (non-parametric Kruskal–Wallis with Dunn's multiple comparisons test). **P < 0.01, ***P < 0.001, ****P < 0.0001. NS, not significant. **d–f,** Synthetic DNA targets containing the exon 16/17 junction sequence were introduced by retroviral transduction into NIH-3T3 cells, and the target sequence (provirus) identified by DISH$_{16/17}$. A concatamer (×2) showed increased focus size, represented as a cumulative frequency distribution plot (**e**) and a box and whisker plot (**f**). Line, median; box, 75th–25th percentiles; whiskers, 90th–10th percentiles. Statistical significance was calculated using non-parametric Kruskal–Wallis test with Dunn's correction for multiple comparisons. ****P < 0.0001.

Extended Data Table 1 | Nine distinct experimental approaches supporting *APP* recombination

	Method	Tested Material	Unit Size of Sample	Reproducibility	Result
1	RT-PCR and Sanger sequencing	Nuclear RNA from human cerebral cortical neurons	50-nuclei	(1) Multiple brains (2) Multiple RT-PCRs (3) Multiple primers (4) Multiple investigators	Novel **APP RNA variants** with IEJs from multiple brains and experiments
2	RISH on IEJ 3/16	Human SAD tissue sections	Tissue sections	(1) Multiple sections (2) Multiple investigators	Cytoplasmic and mosaic **IEJ 3/16** signals on human brain tissue
3	Whole transcriptome SMRT sequencing	Human AD brain RNA prepared by PacBio	Bulk RNA	Multiple scientists using independent approaches identified IEJ variants	Novel **APP RNA variants** with IEJs
4	Targeted RNA SMRT sequencing	Human temporal lobe RNA, pulldown of APP related genes by PacBio	Bulk RNA	Multiple scientists using independent approaches identified IEJ variants	Novel **APP RNA variants** with IEJs
5	PCR and Sanger sequencing	Genomic DNA from SAD and non-diseased cerebral cortical neurons and non-neurons	DNA equivalent to 20-nuclei	(1) Multiple brains (2) Multiple PCRs (3) Multiple primers (4) Multiple polymerases (5) Multiple investigators	Conserved, 8 **gencDNAs** including the same IEJs in genomic DNA as found in RNAs
6	PCR and SMRT sequencing	Genomic DNA from SAD and non-diseased cerebral cortical neurons	Small nuclei populations (20-1000)	(1) Multiple brains (2) Multiple sequencing runs	At least 6,299 unique *APP* **gencDNA** variants from AD brain; 1,084 from non-diseased brain
7	Sequencing of *APP* genomic locus, pulled down by Agilent SureSelect	Genomic DNA from SAD cerebral cortical neurons	200 ng of DNA	Two complete sets of pull-down and sequencing	*APP* **gencDNAs**
8	DISH on gencDNAs	SAD and non-diseased cerebral cortical neuronal nuclei	Single nuclei	(1) Multiple investigators (2) Multiple brains (3) Multiple sense probes	Up to 13 spatially distinct *APP* **gencDNA** loci in single neuronal nucleus, enriched in AD and rare in non-neuronal nuclei
		J20 and WT cerebral cortical neuronal vs. non-neuronal nuclei			Increased *APP* **IEJ 3/16 gencDNA** signal in J20 neurons, but not non-neuronal nuclei
		J20 and WT cerebral cortical neuronal vs. non-neuronal nuclei			Age-related increases in *APP* **Ex 16/17 foci** diameter in J20 neuronal but not non-neuronal nuclei
9	APP751 over-expression in CHO cells with H$_2$O$_2$ treatment	CHO cells	Bulk DNA	(1) Multiple experiments (2) Multiple investigators	*APP* **gencDNA** with IEJs identified, reverse transcriptase activity and DNA breaks dependent

Evidence of *APP* recombination reported in this paper is summarized with experimental methods, materials, unit sizes of sample and reproducibility.

Extended Data Table 2 | Human postmortem brain information

Brain Name	Braak	Sex	PMI (Hours)	Age (years)
SAD-1	6	F	6	88
SAD-2	6	F	12	88
SAD-3	6	F	6	84
SAD-4	6	F	4	86
SAD-5	6	M	5	83
SAD-6	6	F	10	72
SAD-7	5	F	3.7	77
ND-1	1	M	U	87
ND-2	1	F	72	83
ND-3	U	M	U	83
ND-4	1	F	12	80
ND-3	1	F	18	93
ND-6	2	M	12	94

F, female; M, male; U, unknown; PMI, post mortem interval. All brains were from the pre-frontal cortex and obtained from the University of California San Diego Alzheimer's Disease Research Center and the University of California Irvine Institute for Mind Impairments and Neurological Disorders.

Extended Data Table 3 | *APP* variants information

Name	RNA PCR	DNA PCR	Coding or Non-coding	Start (bp)	Break Start	Break End	End	Sanger Sequence Homology	# of bp in homology	# of mis-matches
RT-PCR Identified Variants										
R1_11.1	Y	Y	Non-Coding	1	32	1431	2313	CACTGCTCTGCAGGC	15	3
R1_11.2	Y		Non-Coding	1	44	1456	2313	CGGC	4	0
R1_14	Y	Y	Non-Coding	1	46	1814	2313	AGCTC	5	1
R2_14	Y		Non-Coding	1	200	1749	2313	ACCAAGGA	8	0
R2_16	Y		Non-Coding	1	216	2015	2313	AT	2	0
R2_17	Y	Y	Non-Coding	1	64	2102	2313	CA	2	0
R2_18	Y	Y	Coding	1	211	2267	2313	GC	2	0
R3_14	Y	Y	Coding	1	267	1890	2313	AGCCAAC	7	0
R3_16	Y	Y	Coding	1	251	2008	2313	AA	2	0
R3_17	Y	Y	Non-Coding	1	314	2123	2313	GCAGTG	6	0
R6_17	Y		Coding	1	673	2079	2313	AGATGGGAGTGAAGACAAAG	20	0
R6_18	Y		Coding	1	740	2233	2313	GAGGA	5	0
DNA PCR Identified Variants										
D2_18		Y	Coding	1	120	2287	2310	N/A	N/A	N/A
D1_17		Y	Non-Coding	18	51	2159	2285	N/A	N/A	N/A
D2_16		Y	Non-Coding	19	209	2016	2285	TGCAGAATT	9	1
D2_17		Y	Non-Coding	18	64	2102	2285	CA	2	0
D2_16.2		Y	Non-Coding	157	209	2016	2095	TGCAGAATT	9	1
Commercially available PacBio RNA-Seq										
P3_9	Y	n/a	Non-coding	303	345	1093	+853	CCTAC	5	0
P6_12	Y	n/a	Non-coding	-41	724	1483	+853	AAGAAG	6	0
P6_18	Y	n/a	Non-coding	-111	2029	2274	+936	AATTCCGAC	9	0
DNA PCR on CHO cells: Induced Variants										
iD1_17	n/a	Y	Non-Coding	1	51	2159	2313	N/A	N/A	N/A
iD2_13	n/a	Y	Coding	1	170	1626	2313	N/A	N/A	N/A
iD4_15	n/a	Y	Missense	1	434	1920	2313	TGA	3	1
iD6_18	n/a	Y	Coding	1	705	2269	2313	T	1	0
iD2_17	n/a	Y	Non-Coding	1	64	2102	2313	CA	2	0
RNA PCR on CHO cells: Induced Variants										
iR2/3	Y	n/a	Non-Coding	1	64	319	2313	ACCCA	5	0
iR5/15	Y	n/a	Non-Coding	1	618	1921	2313	GATGACTCG	9	2
iR2/4_7/17	Y	n/a	Non-Coding	1	197/ 934	399/ 2077	2313	CAA/GA	3/2	0/0
iR3/13	Y	n/a	Non-Coding	1	318	1682	2313	AAG	3	0
iR4/18	Y	n/a	Missense	1	397	2285	2313	ACAAGT	6	0

Detailed information on identified *APP* RNA and DNA variants.

Extended Data Table 4 | DISH and RISH experiments and validation list

Junction	Target	Sample	Type	Probes	Figure Panel
16/17	DNA	Human nuclei +RNase	Exp	Sense	Fig. 2d,f,g,n; Fig. 5a,b,c
			Exp	Antisense	Fig. 2d,f,g
		Human nuclei +RNase + restriction enzyme (MluCI)	Neg	Sense	Fig. 2j,k
		Human nuclei +RNase +off-target restriction enzyme (XbaI)	Pos	Sense	ED Fig. 2f
		Synthetic target	Pos	Sense	ED Fig. 2g
			Pos	Antisense	ED Fig. 2g
		Synthetic mutant target	Neg	Sense	ED Fig. 2g
			Neg	Antisense	ED Fig. 2g
		Synthetic target concatamer	Pos	Sense	ED Fig. 6d-f
		WT mouse nuclei +RNase	Neg	Sense	Fig. 5i,j; ED Fig. 6b,c
		J20 mouse nuclei +RNase	Exp	Sense	Fig. 5i,j; ED Fig. 6b,c
3/16	DNA	Human nuclei +RNase	Exp	Sense	Fig. 2h,e,i,n; Fig. 5d,e,f
			Exp	Antisense	Fig. 2h,e,i
		Human nuclei +RNase + restriction enzyme (PSTI & MsII)	Neg	Sense	Fig. 2l,m
		Human nuclei +RNase +off-target restriction enzyme (XbaI)	Pos	Sense	ED Fig. 2f
		Synthetic target	Pos	Sense	ED Fig. 2h
			Pos	Antisense	ED Fig. 2h
		Synthetic mutant target	Neg	Sense	ED Fig. 2h
			Neg	Antisense	ED Fig. 2h
		WT mouse nuclei +RNase	Exp	Sense	Fig. 5g,h; ED Fig. 6a
		J20 mouse nuclei +RNase	Exp	Sense	Fig. 5g,h; ED Fig. 6a
	RNA	SAD tissue	Neg	Sense	ED Fig. 1f
			Exp	Antisense	ED Fig. 1f
			Neg	DapB	ED Fig. 1f
In2/Ex3	DNA	Human nuclei +RNase	Exp	Sense	Fig. 2n

In, intron; Ex, exon; Exp, experimental; Neg, negative control; Pos, positive control; ED, Extended Data. DNA and RNA in situ hybridization experiments, positive and negative controls are summarized.

ARTICLE

Predictable and precise template–free CRISPR editing of pathogenic variants

Max W. Shen[1,2,12], Mandana Arbab[3,4,5,12], Jonathan Y. Hsu[6,7], Daniel Worstell[8], Sannie J. Culbertson[8], Olga Krabbe[8,9], Christopher A. Cassa[8,10], David R. Liu[3,4,5]*, David K. Gifford[2,6,10,11]* & Richard I. Sherwood[8,9]*

Following Cas9 cleavage, DNA repair without a donor template is generally considered stochastic, heterogeneous and impractical beyond gene disruption. Here, we show that template–free Cas9 editing is predictable and capable of precise repair to a predicted genotype, enabling correction of disease-associated mutations in humans. We constructed a library of 2,000 Cas9 guide RNAs paired with DNA target sites and trained inDelphi, a machine learning model that predicts genotypes and frequencies of 1- to 60-base-pair deletions and 1-base-pair insertions with high accuracy ($r = 0.87$) in five human and mouse cell lines. inDelphi predicts that 5–11% of Cas9 guide RNAs targeting the human genome are 'precise–50', yielding a single genotype comprising greater than or equal to 50% of all major editing products. We experimentally confirmed precise–50 insertions and deletions in 195 human disease-relevant alleles, including correction in primary patient–derived fibroblasts of pathogenic alleles to wild-type genotype for Hermansky–Pudlak syndrome and Menkes disease. This study establishes an approach for precise, template–free genome editing.

Clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 has revolutionized genome editing, providing powerful research tools and promising agents for the potential treatment of genetic diseases[1–3]. The DNA-targeting capabilities of Cas9 have been improved by the development of guide RNA (gRNA) design principles[4], modelling of factors leading to off-target DNA cleavage, enhancement of Cas9 sequence fidelity by modifications to the nuclease and gRNA, and the evolution or engineering of Cas9 variants with alternative PAM sequences[5]. Similarly, control over the product distribution of genome editing has been advanced by the development of base editing to achieve precise and efficient single-nucleotide mutations[6,7], and the improvement of template-directed homology-directed repair (HDR) of double-stranded breaks[8]. Despite these developments, base editing does not mediate insertions or deletions, and HDR is limited by low efficiency, particularly in non-dividing cells, and by undesired by-products. As many human genetic variants associated with disease arise from insertions and deletions[9,10], methods to efficiently introduce insertions and deletions to alleviate pathogenic mutations in a predictable manner with a major single-genotype outcome would advance the field of genome editing.

Non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ) processes are major pathways involved in the repair of Cas9-mediated double-stranded breaks that can result in highly heterogeneous repair outcomes comprising hundreds of repair genotypes. Although end-joining repair of Cas9-mediated double-stranded DNA breaks has been harnessed to facilitate knock-in of DNA templates[11,12] or deletion of intervening sequence between two cleavage sites[5], NHEJ and MMEJ are not generally considered useful for precision genome editing applications. Previous work has found that the heterogeneous distribution of Cas9-mediated editing products at a given target site is reproducible and dependent on local sequence context[13,14], but no general methods have been described to predict genotypic products following Cas9-induced double-stranded DNA breaks.

In this study, we developed a high-throughput *Streptococcus pyogenes* Cas9 (*Sp*Cas9)-mediated repair outcome assay to characterize end-joining repair products at Cas9-induced double-stranded breaks using 1,872 target sites based on sequence characteristics of the human genome. We used the resulting rich set of repair product data to train inDelphi, a machine learning algorithm that accurately predicts the frequencies of the substantial majority of template-free Cas9-induced insertion and deletion events at single-base resolution (https://indelphi.giffordlab.mit.edu/). We find that, in contrast to the notion that end-joining repair is heterogeneous, inDelphi identifies that 5–11% of *Sp*Cas9 gRNAs in the human genome induce a single predictable repair genotype in ≥50% of editing products. Building on this idea of precision gRNAs, we used inDelphi to design 14 gRNAs for high-precision template-free editing yielding predictable 1-bp insertion genotypes in endogenous human disease-relevant loci and experimentally confirmed highly precise editing (median 61% among edited products) in two human cell lines. We used inDelphi to reveal human pathogenic alleles that are candidates for efficient and precise template-free gain-of-function genotypic correction and achieved template-free correction of 183 pathogenic human microduplication alleles to the wild-type genotype in ≥50% of all editing products. Finally, we integrate these developments to achieve high-precision correction of five pathogenic low-density lipoprotein receptor (*LDLR*) microduplication alleles in human and mouse cells, as well as correction of endogenous pathogenic microduplication alleles for Hermansky–Pudlak syndrome (*HPS1*) and Menkes disease (*ATP7A*) to the wild-type sequence in primary patient-derived fibroblasts.

[1]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [4]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [5]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA. [6]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [7]Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA, USA. [8]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [9]Hubrecht Institute for Developmental Biology and Stem Cell Research, Royal Netherlands Academy of Arts and Sciences (KNAW), Utrecht, The Netherlands. [10]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [11]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [12]These authors contributed equally: Max W. Shen, Mandana Arbab. *e-mail: drliu@fas.harvard.edu; gifford@mit.edu; rsherwood@rics.bwh.harvard.edu
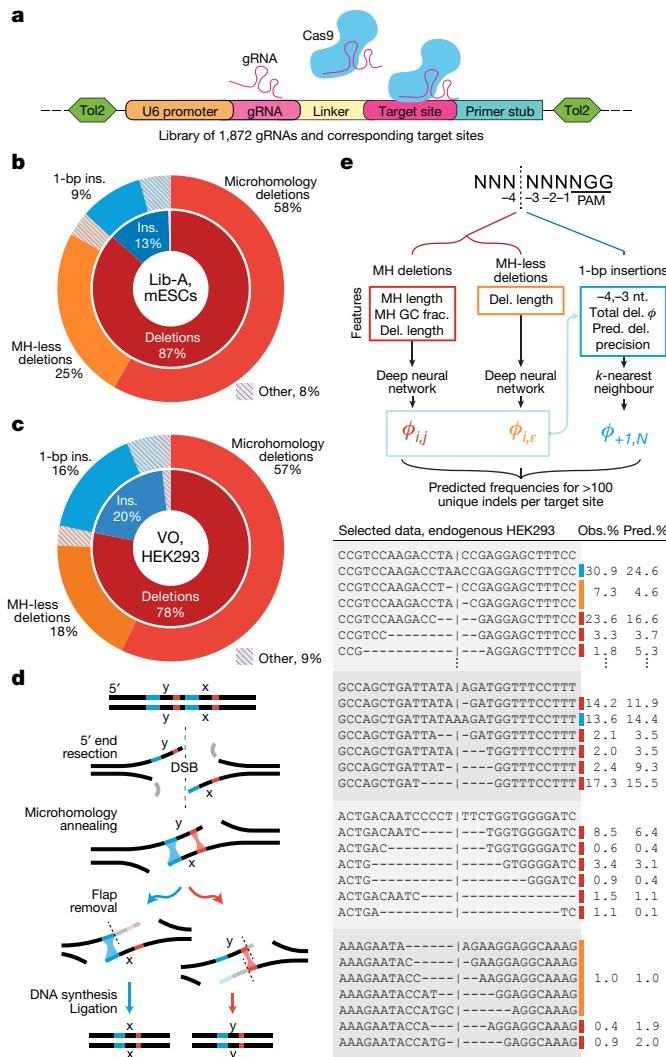
Fig. 1 | High-throughput assaying of Cas9-mediated DNA repair products supports the design of the inDelphi model. a, A high-throughput genome-integrated library for assaying Cas9 editing products. **b**, Categories of editing products at 1,996 lib-A target sites in mESCs. **c**, Categories of editing products in 89 VO endogenous target sites in HEK293 cells. **d**, Mechanism of microhomology-mediated end-joining repair. **e**, inDelphi uses machine learning to predict the frequencies of editing products from target DNA sequence (selected outcomes depicted in table). Major editing outcomes include +1 to −60 indels.

Template–free Cas9 editing is predictable

To capture Cas9-mediated end-joining repair products across a wide variety of target sites, we designed a genome-integrated gRNA and target library screen in which many unique gRNAs are paired with 55-bp target sites containing a single canonical 'NGG' SpCas9 protospacer-adjacent motif (PAM) that directs cleavage to the centre of each target site (Fig. 1a). Previously reported repair products at 90 loci in three human cell lines[14] (HCT116, K562, and HEK293; we refer to the collective data set as VO) showed that 94% of endogenous cut-site proximal Cas9-mediated deletions are ≤30 bp (Extended Data Fig. 1), suggesting that our assay can assess the vast majority of cut-site-proximal editing products. To explore repair products among sequences representative of the human genome, we designed 1,872 target sites spanning the distributions of GC content, number of nucleotides participating in microhomology, predicted Cas9 on-target cutting efficiency[4], and estimated precision of deletions[14] (Supplementary Methods, Extended Data Fig. 1), as well as 90 VO target sites to create a library (denoted hereafter as 'lib-A') in which each target site is accompanied by a corresponding gRNA on the same DNA molecule. Through a multi-step

process (Extended Data Fig. 1), we constructed and cloned lib-A into a plasmid backbone allowing Tol2 transposon-based integration into the genome[15], gRNA expression and hygromycin selection for cells with library members.

We stably integrated lib-A into the genomes of mouse embryonic stem cells (mESCs) and human U2OS cells, then targeted these cells with a Tol2 transposon-based SpCas9 expression plasmid containing a blasticidin expression cassette and selected for cells with stable Cas9 expression while maintaining more than 2,000-fold coverage of the library. After one week, we collected genomic DNA from these cells (three independent biological replicates in mESCs, two in U2OS cells) along with control cells not treated with Cas9 (one in each) and performed paired-end high-throughput DNA sequencing (HTS) to reveal the distribution of cut-site-proximal repair products at each target site (Extended Data Fig. 1). We tabulated the resulting 192,055,534 sequencing reads using a sequence alignment procedure (Supplementary Methods) that identified an average of 245 unique repair outcomes with high confidence (Supplementary Methods) per target site in mESCs (45 in U2OS cells) after adjusting with control data. Repair outcomes in experimental replicates within the same cell type were consistent (median $r = 0.89$ in mESCs, 0.77 in U2OS cells, Extended Data Fig. 1).

In lib-A data from mESCs and U2OS cells as well as in endogenous data in HEK293, K562 and HCT116 cells, end-joining repair of Cas9-mediated double-stranded breaks primarily caused deletions (on average 63–87% of all edited products across cell types) and insertions (13–37% of all products) (Fig. 1b, c, Extended Data Fig. 2). A large fraction of products were deletions containing microhomology consistent with MMEJ (39–58% of all products, 62–75% of deletions, Fig. 1b–d, Extended Data Fig. 2, Supplementary Discussion). Three repair classes constituted 80–95% of all observed editing products (Fig. 1b, c): microhomology (MH) deletions, microhomology-less (MH-less) deletions, and single-base (1-bp) insertions; we define these three repair classes as constituting all major editing outcomes. The insertion and deletion (indel) frequencies at 86 target sites were consistent between endogenous data in HEK293, K562 and HCT116 cells and lib-A data in mESCs and U2OS cells (median $r = 0.65$ to 0.82 for pairs of cell types when adjusting for 1-bp insertion frequencies, median $r = 0.52$ to 0.76 without adjustment, Extended Data Fig. 1). Together, these data confirm that Cas9-mediated editing products from our library assay reflect previously reported endogenous editing in human cells.

Using lib-A, we designed a new machine learning model, inDelphi, to predict the frequency of all major editing outcomes at any given target site. This model consists of three interconnected modules aimed at predicting MH deletions, MH-less deletions and 1-bp insertions (Fig. 1e).

inDelphi predicts MH deletions using a module that simulates the MMEJ repair mechanism, in which 5′–3′ end resection at a double-stranded break reveals two 3′ single-stranded DNA overhangs that can anneal through sequence microhomology. Extraneous single-stranded DNA overhangs are eliminated, and DNA synthesis and ligation generates a double-stranded DNA repair product[16] (Fig. 1d). Through this mechanism, each microhomology results in a distinct deletion genotype (Fig. 1d, Supplementary Discussion). inDelphi assigns a score (phi) to a candidate microhomology based on a neural-network-learned score using its length and GC content with a penalty based on the deletion length. Relative frequencies are obtained by normalizing the phi scores of microhomologies of interest to sum to one, thereby modelling MH deletions as a competitive process.

inDelphi models deletions inconsistent with MMEJ with a second neural network module that predicts the total frequency of groups of MH-less deletion outcomes using the minimum required resection length as the only input feature (Fig. 1e). We hypothesize that MH-less deletions arise primarily from the classical and alternative NHEJ pathways[17] (Supplementary Discussion).

The MH and MH-less neural networks were jointly trained using data from 1,095 lib-A target sites in mESCs with backpropagation in a multi-task manner to predict both deletion length frequencies
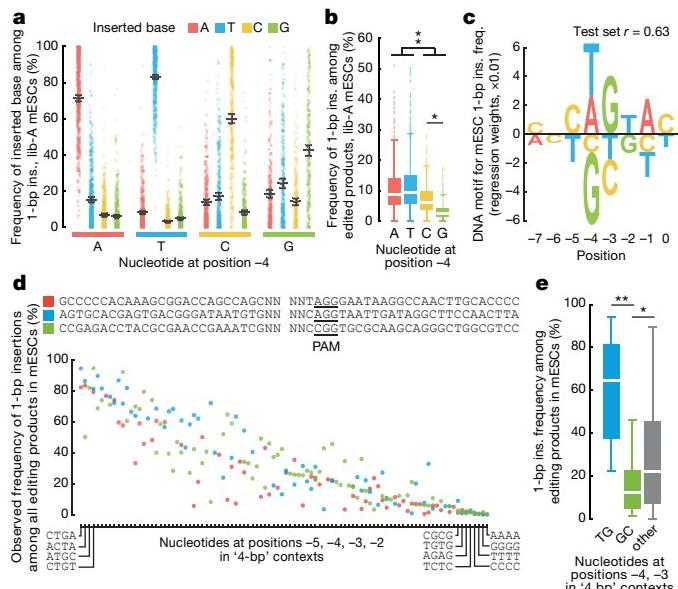
Fig. 2 | Sequence context influences 1-bp insertions. a, 1-bp insertion frequencies (mean ± 95% confidence intervals) among 1,981 lib-A target sites. **b**, Comparison of 1-bp insertion frequencies among Cas9-edited products from 1,996 lib-A target sites. The box denotes the 25th, 50th and 75th percentiles, whiskers show 1.5 times the interquartile range, and outliers are depicted as diamonds. $*P = 5.4 \times 10^{-36}$, $**P = 8.6 \times 10^{-70}$, two-sided t-test. **c**, DNA motif for 1-bp insertion frequency (lib-A, mESCs, $n = 1,996$ target sites). **d**, Frequencies of 1-bp insertions among 205 target sites with varying −5 to −2 nucleotides (relative to the PAM at positions 0–2) in three low-microhomology contexts. See Extended Data Fig. 5 for full axis labels. **e**, Comparison of the 1-bp insertion frequency at sequences in **c** with varying positions −4 and −3. Box plot as in **b**. $*P = 0.03$, $**P = 2.98 \times 10^{-7}$, two-sided t-test.

and MH genotype frequencies (Fig. 1e, Supplementary Methods). Computational experiments confirmed that the design of the neural network modules was important for overall performance (Supplementary Methods). From training data, inDelphi learned that strong microhomologies tend to be long and have high GC content and that the frequency of MH-less deletions decays rapidly with increasing length (Extended Data Fig. 2). For 1- to 30-bp deletions, at a typical target site in the human genome, inDelphi makes one prediction for each of 92 possible MH deletions, and 30 predictions for 274 possible MH-less deletion genotypes.

inDelphi contains a third module that uses k-nearest neighbours to predict 1-bp insertions (Fig. 1e), which represent a major class of edited products (9–30% of all edited products, Fig. 1b, Extended Data Fig. 2). The frequency of 1-bp insertions and their resultant genotypes depend strongly on local sequence context. They are predominantly duplications of the −4 nucleotide (counting the NGG PAM as nucleotides 0–2, Fig. 1e), with higher precision and frequency when the −4 nucleotide is an A or T (Fig. 2a, b). A linear regression model trained to predict the frequency of 1-bp insertions among major editing outcomes from local sequence context performed well on held-out lib-A target sites (that is, those not included in the training of the model) in mESCs ($n = 499$, $r = 0.63$, Fig. 2c) and U2OS cells ($n = 492$, $r = 0.65$, Extended Data Fig. 3). In both cell types, target sites with weak microhomology (low total phi score) or low deletion precision score (Supplementary Methods) were significantly more likely to yield insertions at the expense of deletions ($P < 2.0 \times 10^{-3}$, Extended Data Fig. 3). Randomization of four nucleotides surrounding the Cas9 cleavage site in three constant background sequences with weak microhomology revealed substantial variation in 1-bp insertion frequency (from ≤5% to ≥80% of all edited products, Fig. 2d, Extended Data Fig. 3) and identified mini-motifs consistent with lib-A (Fig. 2e), suggesting that local sequence context is a highly influential and causal factor for 1-bp insertion repair.

On the basis of these data, inDelphi models insertions and deletions as competitive processes in which microhomology strength and precision of deletions influence the relative frequency of 1-bp insertions, and local sequence context influences the relative frequency and genotypic outcomes of 1-bp insertions (Fig. 1e). inDelphi makes predictions within each module in a cell-type-invariant manner, only using cell-type-specific data to predict the overall ratio of 1-bp insertions to deletions. Collectively across all three modules, inDelphi predicts the indel lengths of 80–95% of Cas9-mediated editing products and the genotypes of 65–80% of all products (Fig. 3a, Extended Data Fig. 4) from sequence context alone.

inDelphi achieves high accuracy in predicting genotype frequencies (median $r = 0.94$) and indel length frequency distributions (median $r = 0.91$) in 189 held-out lib-A target sites in mESCs (Extended Data Fig. 4), with similarly high accuracy in U2OS cells (median $r = 0.88$ and 0.91, Extended Data Fig. 4). On held-out endogenous data, inDelphi also performed strongly on the two tasks (median $r = 0.87$ and 0.84 across 87–90 target sites in K562, HCT116 and HEK293 cells, Fig. 3b, c). Taken together, these results establish that in data from five human and mouse cell lines, the relative frequencies of most Cas9-nuclease-mediated editing outcomes are highly predictable.

The ability of Cas9-mediated end-joining repair to induce frameshifts enables efficient gene knockout[5]. We reasoned that inDelphi's accurate prediction of indel lengths when considering nearly all editing products would enable accurate prediction of Cas9-induced frameshifts. We simulated this task in data from 82–91 endogenous target sites by tabulating the observed frequency of indels resulting in +0, +1 and +2 reading frames. In HEK293 cells, the observed frequency of indels in each frame predicted by inDelphi (median $r = 0.81$) compare favourably to those generated by 'Microhomology Predictor', a previously published method[18] (median $r = 0.37$, Fig. 3d), with similar results in HCT116 and K562 cells (Extended Data Fig. 4). Thus, we expect inDelphi to facilitate Cas9-mediated gene knockout approaches by allowing a priori selection of gRNAs that induce high or low knockout frequencies. We note that microhomology deletions in human exons have a significant tendency to remain in-frame compared to non-coding human DNA (Extended Data Fig. 4).

Highly precise template–free Cas9 editing

Although end-joining repair is highly efficient in inducing mutations after Cas9 treatment, its propensity to induce a heterogeneous mixture of repair genotypes has limited applications for precision genome editing[19]. We used inDelphi to estimate the fraction of SpCas9 gRNAs targeting exons and introns in the human genome that support precise end-joining repair. Defining 'precision-x' gRNAs as those predicted to produce a single genotypic outcome in ≥x% of all major editing outcomes proximal to the cleavage site, inDelphi predicts that 28% and 47% of gRNAs are precision-30, whereas 5% and 11% of gRNAs are precision-50, when trained on mESC and U2OS cell data, respectively (Fig. 3f, Extended Data Table 1).

To test the accuracy of the predictions made using inDelphi of precise repair in endogenous settings, we selected 14 SpCas9 gRNAs predicted to induce precision-40 1-bp insertions. We delivered SpCas9 with gRNAs and performed endogenous HTS in human U2OS and HEK293T cells. We observed that 10 out of 14 predicted precision-40 1-bp insertion gRNAs induced a single 1-bp insertion genotype in ≥40% of edited products with an overall significantly higher precision ($P < 4.2 \times 10^{-8}$) than baseline data in HEK293T (median 55% compared with 25% baseline in VO target sites in HEK293) and U2OS cells (median 57% compared with 14% baseline in lib-A, U2OS, Fig. 3e). We similarly validated 10 gRNAs for high-precision deletions with endogenous HTS in both cell types (Extended Data Table 2). Collectively, these observations establish the ability of inDelphi to identify, from sequence features alone, gRNAs that induce significantly more precise editing than the general population of gRNAs.
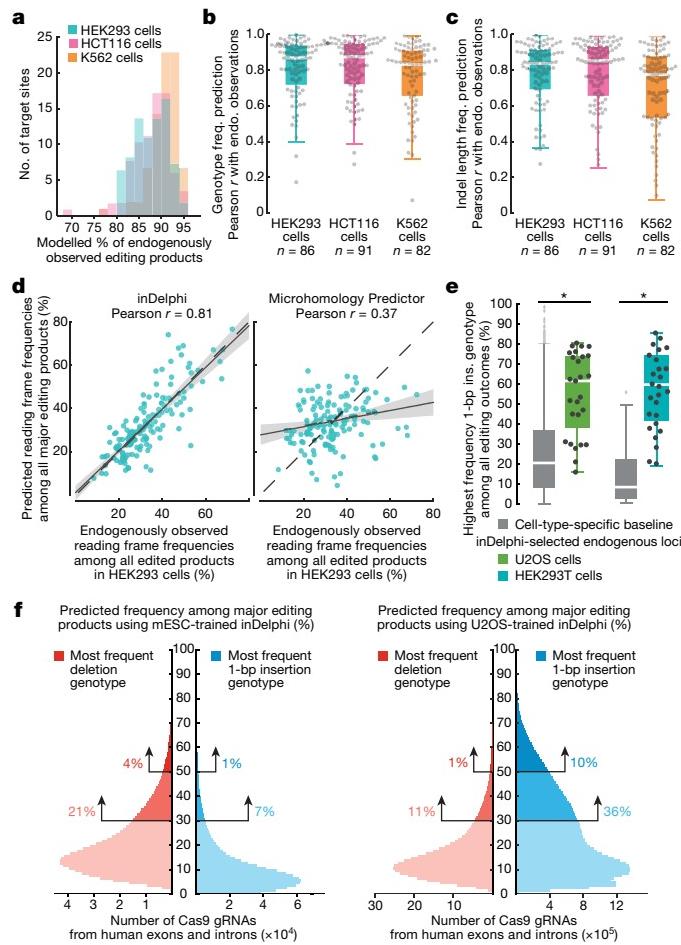
Fig. 3 | inDelphi accurately predicts nearly all editing outcomes.
a, Fraction of endogenous editing products given predictions in HEK293
($n = 86$ target sites), HCT116 ($n = 91$) and K562 cells ($n = 82$).
b, **c**, Predictive performance on endogenously observed frequencies of
genotypes (**b**) and indel lengths (**c**) in HEK293 (median, 0.87 and 0.84),
HCT116 (median, 0.87 and 0.85), and K562 (median, 0.83 and 0.79) cells.
The box denotes the 25th, 50th and 75th percentiles, and whiskers show
1.5 times the interquartile range. **d**, Comparison of predictions from two
methods to observed frame frequencies ($n = 86$ target sites, HEK293
cells), regression estimate $\pm$ 95% confidence intervals. **e**, 1-bp insertion
frequencies among edited outcomes in U2OS and HEK293T cells ($n = 27$
and 26 observations, baseline $n = 1,958$ and 89 target sites, $P = 4.2 \times 10^{-8}$
and 8.1×10^{-12}, respectively), two-sided Welch's t-test. **f**, Smoothed
predicted distribution of the highest frequency indel among major editing
outcomes ($+1$ to -60 indels) for SpCas9 gRNAs targeting the human genome.

Template–free correction of pathogenic alleles

We used inDelphi to identify new targets for therapeutic genome edit-
ing. Starting with 23,018 pathogenic short indels (ClinVar and HGMD
databases[9,10]), we used inDelphi to identify pathogenic alleles that are
suitable for template-free Cas9-mediated editing to effect precise gain-
of-function editing of the pathogenic genotype. We pursued two genetic
disease allele categories that have not been previously identified as targets
for Cas9-mediated repair: pathogenic frameshifts in which inDelphi pre-
dicts that 50–90% of Cas9-mediated deletion products will correct the
reading frame (mean baseline frequency of 34% among disease-associated
frameshift mutations) and pathogenic microduplication alleles in which
a short sequence duplication leads to a frameshift mutation or disrupts
protein function and which inDelphi predicts can be repaired to wild-type
genotype in a large fraction of Cas9 editing products (Fig. 4a).

We selected 1,592 pathogenic human loci with high predicted rates
of frame correction or microduplication correction to the wild-type
sequence for inclusion in a second library (lib-B). We observed that
183 human disease microduplication alleles included in lib-B were

repaired to wild-type in $\geq$50% of all products (Fig. 4b), and 508 path-
ogenic human frameshift alleles were corrected into proper reading
frames in $\geq$50% of all products in mESCs (Fig. 4c), in agreement with
inDelphi's predictions ($r = 0.64$ and 0.64). We observed similar results
in U2OS cells ($r = 0.65$ for frame correction, $r = 0.61$ for genotype cor-
rection to wild type, Extended Data Fig. 5). Although repair to the
wild-type genotype unambiguously restores wild-type protein function,
we note that frame correction that alters the coding sequence requires
case-by-case analysis to validate rescue of protein function.

To determine whether the efficiency of microduplication repair can be
increased by manipulation of DNA repair pathways, we performed Cas9
cleavage of lib-B in four NHEJ-deficient conditions[20]: $Prkdc^{-/-}Lig4^{-/-}$
mESCs[21], and mESCs treated separately with DNA-dependent protein
kinase inhibitor III (DPKi3), NU7026 and MLN4924. In NHEJ-impaired
cells, the fraction of deletion outcomes not involving MH significantly
decreased (median 23% to 10% with $Prkdc^{-/-}Lig4^{-/-}$, $P = 1.0 \times 10^{-36}$,
and 23% to 19% with DPKi3 and NU7041, $P < 5.5 \times 10^{-5}$) (Extended
Data Fig. 6, Supplementary Discussion). In $Prkdc^{-/-}Lig4^{-/-}$ mESCs, the
increased propensity towards MH deletions enabled a subset of patho-
genic alleles to be repaired to wild type with markedly high precision.
Compared to wild-type mESCs in which 183 pathogenic alleles cor-
rected to wild type in $\geq$50% of all edited products and 11 pathogenic
alleles corrected to wild type in $\geq$70% of all edited products, in $Prkdc^{-/-}$
$Lig4^{-/-}$ mESCs, 286 pathogenic alleles corrected to wild type in $\geq$50%
of all edited products and 153 pathogenic alleles corrected to wild type
in $\geq$70% of products (Fig. 4d, Supplementary Table 1) without increase
in the rate of apoptosis (Extended Data Fig. 6). DPKi3 or NU7041
treatment also increased precise microduplication repair (Extended
Data Figs. 5, 6). Taken together, impairing NHEJ can further increase
the precision of wild-type correction for a large subset of pathogenic
microduplications in genes such as $PKD1$ (corrected in 92% of edited
$Prkdc^{-/-}Lig4^{-/-}$ mESC alleles), $MSH2$ (88%) and $LDLR$ (87%), supporting
a model of competing end-joining repair mechanisms.

We further tested inDelphi's prediction of highly efficient correction
in a functional assay with pathogenic $LDLR$ microduplication alleles
that cause dominantly inherited familial hypercholesterolemia[22]. We
separately introduced five pathogenic $LDLR$ microduplication alleles
within a full-length $LDLR$ coding sequence upstream of a P2A–
GFP cassette into the genome of human and mouse cells, such that
Cas9-mediated repair to the wild-type LDLR sequence should induce
phenotypic gain of LDL uptake and restore the reading frame of GFP. We
then delivered Cas9 and a gRNA that is specific to each pathogenic allele
and does not target the wild-type repaired sequence. We observed robust
restoration of LDL uptake as well as restoration of GFP fluorescence in
mESCs, U2OS cells and HCT116 cells in up to 79% of cells following
transfection with Cas9 and inDelphi gRNAs (Fig. 4e, f, Extended Data
Fig. 7). HTS confirmed efficient correction of these five LDLR micro-
duplication alleles to wild type in human and mouse cells, as well as path-
ogenic microduplication alleles in the GAA, $GLB1$ and $PORCN$ genes
introduced to cells using the same method (Extended Data Table 3).
Importantly, in these experiments, we observed high-frequency LDLR
phenotypic correction when cutting with either SpCas9 or $Streptococcus$
$aureus$ Cas9 (SaCas9)[23] (Extended Data Table 3).

Finally, we used precise template-free Cas9-mediated MMEJ to
correct pathogenic microduplication alleles endogenously in patient-
derived fibroblasts for Hermansky–Pudlak syndrome ($HPS1$ gene),
which causes blood clotting deficiency and albinism in patients and is
particularly common in Puerto Ricans[24], and Menkes disease ($ATP7A$
gene), which results in copper deficiency. Simultaneous delivery of
Cas9 and gRNA specific to the pathogenic microduplication allele
induced high-efficiency correction to the wild-type sequence in $HPS1$
(mean frequency, 88% of edited alleles, $n = 5$ independent biological
experiments) and $ATP7A$ (frequency, 94% of edited alleles, $n = 2$,
Extended Data Table 2). These findings demonstrate the potential of
template-free, precise Cas9-nuclease-mediated repair of microdupli-
cation alleles to achieve efficient repair to the wild-type sequence for
therapeutic gain-of-function genome editing.
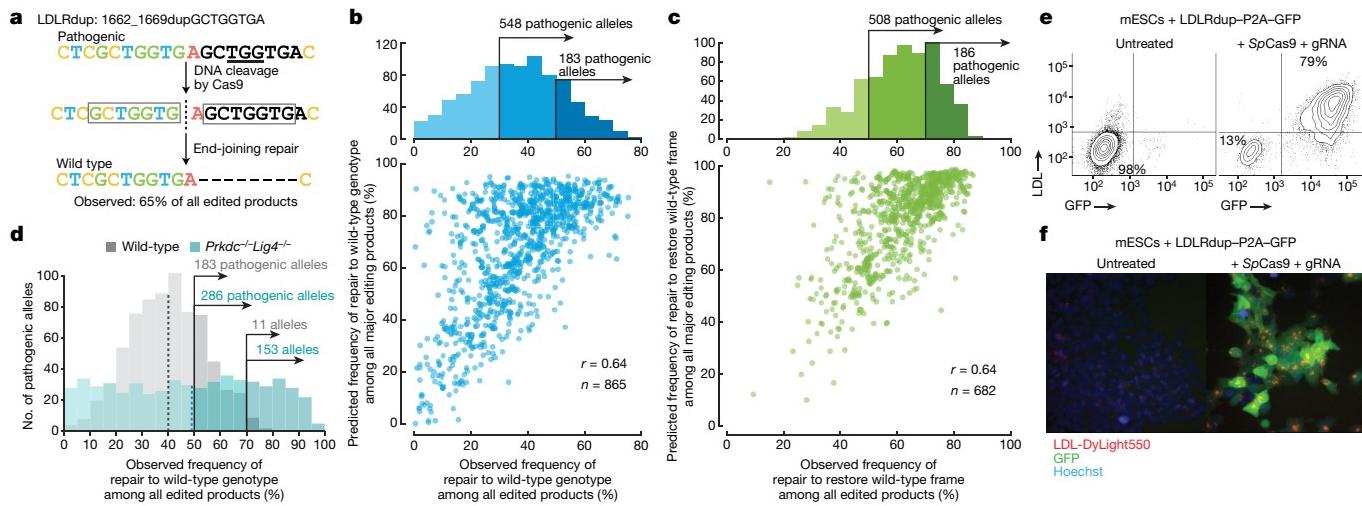
Fig. 4 | Precise template-free correction of pathogenic alleles.
a, Efficient correction of a pathogenic allele to wild type. **b**, **c**, Comparison among pathogenic alleles of observed and predicted frequencies of repair to wild-type genotype (**b**) and frame (**c**). **d**, Wild-type repair frequencies of pathogenic alleles with predicted frequency ≥50% among all major editing outcomes in mESCs. Dashed lines indicate means. **e**, **f**, For mESCs containing the LDLRdup$^{1662\_1669dupGCTGGTGA}$–P2A–GFP allele, flow cytometric contour plots (**e**) and fluorescence microscopy (**f**). Representative data for $n = 2$ independent biological replicates. Major editing outcomes include +1 to −60 indels.

Discussion

We used the Cas9-mediated end-joining repair products of thousands of target DNA loci integrated into mammalian cells to train a machine learning model, inDelphi, that accurately predicts the spectrum of cut-site proximal genotypic products resulting from double-stranded break repair at a target DNA site of interest. The ability to predict Cas9-mediated products enables new precision genome editing research applications and facilitates existing applications, such as performing efficient bi-allelic gene knockout and predicting end-joining by-products of HDR. We provide an online implementation of inDelphi to predict the spectrum of Cas9-mediated products along with predicted frameshift frequencies and precision at any target site (https://indelphi.giffordlab.mit.edu/).

The inDelphi model identifies target loci in which a substantial fraction of all repair products consist of a single genotype. Our findings suggest that 28–47% of SpCas9 gRNAs that target the human genome yield a single indel genotype in ≥30% of all major repair products (precision-30), and 5–11% yield a single indel genotype in ≥50% of all major repair products (precision-50). We show experimentally that precision template-free Cas9-mediated editing can mediate efficient gain-of-function repair at hundreds of pathogenic alleles including microduplications (Fig. 4b, e, f) in cell lines and in patient-derived primary cells (Extended Data Table 3). We note that each research or therapeutic Cas9-nuclease application may require a different level of precision depending on a variety of factors including risk/reward calculations of the gene and disease in question.

Moreover, we present evidence that suppressing NHEJ augments repair of pathogenic microduplication alleles, suggesting that temporary manipulation of DNA repair pathways could be combined with Cas9-mediated editing to favour specific editing genotypes with high precision. Genome editing currently lacks flexible strategies to correct indels in post-mitotic cells because of the limited efficiency of HDR in non-dividing cells[19]. As MMEJ is thought to occur throughout the cell cycle[25], inDelphi may provide access to predictable and precise post-mitotic genome editing in a wider range of cell states. Incorporating the frequencies of long deletions and translocations[26,27] into predictive models of Cas9 outcomes will be an important next step to calculate the overall precision of Cas9-nuclease editing. We anticipate that, given appropriate training data, inDelphi will also be able to accurately predict repair genotypes from other designer nucleases[5]. This work establishes that the prediction and judicious application of template-free Cas9 nuclease-mediated genome editing offers new capabilities for the study and potential treatment of genetic diseases.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0686-x.

1. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
2. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
3. Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
4. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
5. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).
6. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
7. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
8. Paquet, D. et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).
9. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
10. Stenson, P. D. et al. Human Gene Mutation Database: towards a comprehensive central mutation database. *J. Med. Genet.* **45**, 124–126 (2008).
11. Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
12. Nakade, S. et al. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* **5**, 5560 (2014).
13. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, Mdel. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
14. van Overbeek, M. et al. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol. Cell* **63**, 633–646 (2016).
15. Urasaki, A., Morvan, G. & Kawakami, K. Functional dissection of the Tol2 transposable element identified the minimal *cis*-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* **174**, 639–649 (2006).
16. Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* **26**, 52–64 (2016).
17. Deriano, L. & Roth, D. B. Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu. Rev. Genet.* **47**, 433–455 (2013).
18. Bae, S., Kweon, J., Kim, H. S. & Kim, J.-S. Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* **11**, 705–706 (2014).
19. Cornu, T. I., Mussolino, C. & Cathomen, T. Refining strategies to translate genome editing to the clinic. *Nat. Med.* **23**, 415–423 (2017).

20. Davis, A. J. & Chen, D. J. DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.* **2**, 130–143 (2013).
21. Arbab, M., Srinivasan, S., Hashimoto, T., Geijsen, N. & Sherwood, R. I. Cloning-free CRISPR. *Stem Cell Reports* **5**, 908–917 (2015).
22. Bourbon, M., Alves, A. C. & Sijbrands, E. J. Low-density lipoprotein receptor mutational analysis in diagnosis of familial hypercholesterolemia. *Curr. Opin. Lipidol.* **28**, 120–129 (2017).
23. Ran, F. A. et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
24. Oh, J. et al. Positional cloning of a gene for Hermansky–Pudlak syndrome, a disorder of cytoplasmic organelles. *Nat. Genet.* **14**, 300–306 (1996).
25. Biehs, R. et al. DNA double-strand break resection occurs during non-homologous end joining in G1 but is distinct from resection during homologous recombination. *Mol. Cell* **65**, 671–684 (2017).
26. Shin, H. Y. et al. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat. Commun.* **8**, 15464 (2017).
27. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0686-x.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0686-x.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.R.L. or D.K.G. or R.I.S.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Library cloning. In brief, the cloning process involves ordering a library of oligonucleotides pairing a gRNA protospacer with its 55-bp target site, centred on an NGG PAM. To insert the gRNA hairpin between the gRNA protospacer and the target site, the library undergoes an intermediate Gibson Assembly circularization step, restriction enzyme linearization and Gibson Assembly into a plasmid backbone containing a U6 promoter to facilitate gRNA expression, a hygromycin-resistance cassette and flanking Tol2 transposon sites to facilitate integration into the genome.

Specified pools of 2,000 oligonucleotides were synthesized by Twist Bioscience and amplified with NEBNext polymerase (New England Biolabs) using primers 'oligonucleotide library forward' and 'oligonucleotide library reverse' (see below), to extend the sequences with overhangs complementary to the donor template used for circular assembly. To avoid overamplification in the library cloning process, we first performed qPCR by addition of SybrGreen Dye (Thermo Fisher) to determine the number of cycles required to complete the exponential phase of amplification. We ran the PCR reaction for half of the determined number of cycles at this stage. Extension time for all PCR reactions was extended to 1 min per cycle to prevent skewing towards GC-rich sequences. The 246-bp fragment was purified using a PCR purification kit (Qiagen).

Separately, the donor template for circular assembly was amplified with NEBNext polymerase for 20 cycles from an *Sp*Cas9 sgRNA expression plasmid (Addgene 71485)[21] using primers 'circular donor forward' and 'circular donor reverse' (see below) to amplify the sgRNA hairpin and terminator, and extended further with a linker region meant to separate the gRNA expression cassette from the target site in the final library. The 146-bp amplicon was gel-purified (Qiagen) from a 2.5% agarose gel.

The amplified synthetic library and donor templates were ligated by Gibson Assembly (New England Biolabs) in a 1:3 molar ratio for 1 h at 50 °C, and unligated fragments were digested with Plasmid Safe ATP-Dependent DNase (Lucigen) for 1 h at 37 °C. Assembled circularized sequences were purified using a PCR purification kit (Qiagen), linearized by digestion with SspI (New England Biolabs) for ≥3 h at 37 °C, and the 237-bp product was gel-purified (Qiagen) from a 2.5% agarose gel.

The linearized fragment was further amplified with NEBNext polymerase using primers 'plasmid insert forward' and 'plasmid insert reverse' (see below) for the addition of overhangs complementary to the 5′ and 3′ regions of a Tol2 transposon containing gRNA expression plasmid (Addgene 71485)[21] previously digested with BbsI and XbaI (New England Biolabs), to facilitate gRNA expression and integration of the library into the genome of mammalian cells. To avoid overamplification, we performed qPCR by addition of SybrGreen Dye (Thermo Fisher) to determine the number of cycles required to complete the exponential phase of amplification, and then ran the PCR reaction for the determined number of cycles. The 375-bp amplicon was gel-purified (Qiagen) from a 2.5% agarose gel.

The 375-bp amplicon and double-digested Tol2 transposon containing gRNA expression plasmid (Addgene 71485)[21] were ligated by Gibson Assembly (New England Biolabs) in a 3:1 ratio for 1 h at 50 °C. Assembled plasmids were purified by isopropanol precipitation with GlycoBlue Coprecipitant (Thermo Fisher) and reconstituted in milliQ water and transformed into NEB10beta (New England Biolabs) electrocompetent cells. Following recovery, a small dilution series was plated to assess transformation efficiency and the remainder was grown in liquid culture in DRM medium overnight at 37 °C. A detailed step-by-step library cloning protocol is provided in the Supplementary Methods.

The plasmid library was isolated by Midiprep plasmid purification (Qiagen). Library integrity was verified by restriction digest with SapI (New England Biolabs) for 1 h at 37 °C, and sequence diversity was validated by HTS as described below.

Library cloning primers. Oligonucleotide library forward: 5′-TTTT TGTTTTCTGTGTTCCGTTGTCCGTGCTGTAACGAAAGGATGGGTGCGAC GCGTCAT-3′; oligonucleotide library reverse: 5′-GTTGATAACGGACTAGCC TTATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC-3′; circular donor forward: 5′-GTTTAAGAGCTATGCTGGAAACAGC-3′; circular donor reverse: 5′-ATGACGCGTCGCACCCATCCTTTCGTTACAGC ACGGACAACGGAACACAGAAAACAAAAAAGCACCGACTC-3′; plasmid insert forward: 5′-GTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATC TTGTGGAAAGGACGAAACACC-3′; plasmid insert reverse: 5′-TTGTGGTTT GTCCAAACTCATCAATGTATCTTATCATGTCTGCTCGAAGCGGCCGTACC TCTAGAT TCAGACGTGTGCTCTTCCGATCT-3′.

Cloning. A base plasmid was constructed starting from a Tol2-transposon-containing plasmid (Addgene 71485)[21]. The sequence between Tol2 sites was replaced with a CAGGS promoter, multi-cloning site, P2A peptide sequence followed eGFP sequence, and puromycin-resistance cassette to produce p2T-CAG-MCS-P2A-GFP-PuroR. The full sequence of this plasmid is appended in

the 'Sequences' section of the Supplementary Methods, and this plasmid has been submitted to Addgene as catalogue number 107186. Plasmids with this backbone and containing wild-type and microduplication mutation versions of *LDLR* and three other genes, *GAA*, *GLB1*, and *PORCN*, were constructed. Information on cloning these genes is provided below, and the gene sequences are appended in the Supplementary Methods.

LDLR: To generate p2T-CAGGS-LDLRwt-P2A-GFP-PuroR, LDLR (NCBI gene ID 3949, transcript variant 1 CDS) was PCR-amplified from a base plasmid ordered from the Harvard PlasmID resource core and cloned between the BamHI and NheI sites of the base plasmid. The following mutants were generated through InFusion (Clontech) cloning. Sequences are provided below, and our internal allele nomenclature is in parentheses: *LDLR*: c.526_533dupGGCTCGGA (LDLRdup252); *LDLR*: c.668_681dupAGGACAAATCTGAC (LDLRdup254/255); *LDLR*: c.669_680dupGGACAAATCTGA (LDLRdup258); *LDLR*: c.672_683dup-CAAATCTGACGA (LDLRdup261); *LDLR*: c.1662_1669dupGCTGGTGA (LDLRdup264).

PORCN: NCBI gene ID 64840, transcript variant C CDS was PCR-amplified from HCT116 cDNA and cloned between the BamHI and NheI sites of the base plasmid. *PORCN*: c.1059_1071dupCCTGGCTTTTATC (PORCNdup20) was generated through InFusion cloning.

GLB1: NCBI gene ID 2720, transcript variant 1 CDS was PCR-amplified from HCT116 cDNA and cloned between the BamHI and NheI sites of the base plasmid. *GLB1*: c.1456_1466dupGGTGCATATAT (GLB1dup84) was generated through InFusion cloning.

GAA: NCBI gene ID 2548, transcript variant 1 CDS was PCR-amplified from a base plasmid ordered from the Harvard PlasmID resource core and cloned between the BamHI and NheI sites of the base plasmid. *GAA*: c.2704_2716dupCAGAAG-GTGACTG (GAAdup327/328) was generated through InFusion cloning.

*Sp*Cas9[1]: CDS was amplified from p2T-CAG-SpCas9-BlastR and cloned between the BamHI and NheI sites of the base plasmid by Gibson Assembly.

*Sp*Cas9[1] and KKH *Sa*Cas9[28] were constructed starting from a Tol2-transposon-containing plasmid (Addgene 71485)[21]. The sequence between Tol2 sites was replaced with a CAGGS promoter, Cas9 sequence, and blasticidin-resistance cassette to produce p2T-CAG-SpCas9-BlastR and p2T-CAG-KKHSaCas9-BlastR. These plasmids have been submitted to Addgene as catalogue numbers 107189 and 107190.

*Sp*Cas9 gRNAs were cloned as a pool into a Tol2-transposon-containing gRNA expression plasmid (Addgene 71485)[21] using BbsI plasmid digest and Gibson Assembly (New England Biolabs). *Sa*Cas9 gRNAs were cloned into a similar Tol2-transposon-containing *Sa*Cas9 gRNA expression plasmid (p2T-U6-sgSaCas2 x BbsI-HygR), which has been submitted to Addgene using BbsI plasmid digest and Gibson Assembly. Protospacer sequences used are listed below, using our internal nomenclature which matches the duplication alleles.

LDLR gRNAs: sgsaLDLRdup252, 5′-GCTGCGAAGATGGCTCGGAGGC-3′; sgsaLDLRdup254, 5′-GTGCAAGGACAAATCTGACAGG-3′; sgsaLDLRdup255, 5′-GTTCCTCGTCAGATTTGTCCTG-3′; sgsaLDLRdup258, 5′-GACTG CAAGGACAAATCTGAGG-3′; sgsaLDLRdup261, 5′-GTTTTCCTCGTC AGATTTGTCG-3′; sgspLDLRdup264, 5′-GACATCTACTCGCTGGTGAGC-3′. *PORCN* gRNAs: sgspPORCNdup20, 5′-GCTGTCCCTGGCTTTTATCCC-3′. *GLB1* gRNAs: sgspGLB1dup84, 5′-GTGTGAACTATGGTGCATATA-3′. *GAA* gRNAs: sgsaGAAdup327, 5′-GCAGCTGCAGAAGGTGACTGCA-3′; sgsp-GAAdup328, 5′-GCTGCAGAAGGTGACTGCAGA-3′.

Cell culture. mESC lines used have been described previously and were cultured as described previously[29]. HEK293T, HCT116, and U2OS cells were purchased from ATCC and cultured as recommended by ATCC. The following cell lines were obtained and cultured as recommended from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research: GM14609 Hermansky–Pudlak Syndrome 1 (HPS1) fibroblasts and GM13672 Menkes Syndrome fibroblasts. Cell lines were authenticated by the suppliers and tested negative for mycoplasma.

For stable Tol2 transposon plasmid integration, cells were transfected using Lipofectamine 3000 (Thermo Fisher) following standard protocols with equimolar amounts of Tol2 transposase plasmid[15] (a gift from K. Kawakami) and transposon-containing plasmid. For library applications, 15-cm plates with >10^7 initial cells were used, and for single gRNA targeting, 6-well plates with >10^6 initial cells were used. To generate lines with stable Tol2-mediated genomic integration, selection with the appropriate selection agent at an empirically defined concentration (blasticidin, hygromycin, or puromycin) was performed starting 24 h after transection and continuing for >1 week. In cases where sequential plasmid integration was performed such as integrating gRNA/target library and then Cas9 or microduplication plasmid and then Cas9 plus gRNA, the same Lipofectamine 3000 transfection protocol with Tol2 transposase plasmid was performed each time, and >1 week of appropriate drug selection was performed after each transfection.

For *Sp*Cas9 targeting experiments, cells were transduced with a single lentivirus containing an *Sp*Cas9 and sgRNA expression cassette to target *Sp*Cas9 cleavage to either the *HPS1*: c.1472_1487dup16 or *ATP7A*: c.6913_6917dupCTTAT microduplication locus for use in HPS1 and Menkes Syndrome fibroblasts, respectively. The lentiviral plasmids were obtained from (LV01, Sigma-Aldrich) and lentivirus was produced by the Boston Children's Hospital Viral Core. Fibroblasts were plated in 12-well plates at 12,500 cells cm^{-2} one day before transduction. Cells were treated with 10–20 μl of virus in the presence of 8 μg ml^{-1} Polybrene (Sigma-Aldrich) on two consecutive days and collected on day 10 after transduction.

Apoptosis analysis. Wild-type and *Prkdc*$^{-/-}$*Lig4*$^{-/-}$ mESCs with stably integrated Lib-A were transfected with p2T-CAG-SpCas9-P2A-GFP-PuroR using Lipofectamine 3000 following standard protocols in 6-well plates with 10^6 cells. After 24 h, cells were stained with Annexin V Alexa Fluor 568 conjugate (Thermo Fisher) according to the manufacturer's protocols. Fluorescence was detected on a Cytoflex LX (Beckman Coulter) and analysed using FlowJo (FlowJo LLC).

Deep sequencing. Genomic DNA was collected from cells after >1 week of selection. For library samples, 16 μg gDNA was used for each sample; for individual locus samples, 2 μg gDNA was used; for plasmid library verification, 0.5 μg purified plasmid DNA was used.

For individual locus samples, the locus surrounding CRISPR–Cas9 mutation was PCR-amplified in two steps using primers >50-bp from the Cas9 target site. PCR1 was performed using the primers specified below. PCR2 was performed to add full-length Illumina sequencing adapters using the NEBNext Index Primer Sets 1 and 2 (New England Biolabs) or internally ordered primers with equivalent sequences. All PCRs were performed using NEBNext polymerase. Extension time for all PCR reactions was extended to 1 min per cycle to prevent skewing towards GC-rich sequences. The pooled samples were sequenced using NextSeq (Illumina) at the Harvard Medical School Biopolymers Facility, the MIT BioMicro Center, or the Broad Institute Sequencing Facility. The library prep primers were as follows. For LDLRDup252, 254, 255, 258, 261:120417_LDLRDup254_r1seq_A, 5′-CTTTCCCTACACGACGCTCTTCCGATCT NNN ACTCCAGCTGGCGCTGTGAT-3′; 120417_LDLR254_r2seq_A, 5′-GGA GTTCAGACGTGTGCTCTTCCGATCTCAACTTCATCGCTCATGTCCTTG-3′. For LDLRDup264: 120817_LDLR264_r1seq_B, 5′-CTTTCCCTACACGA CGCTCTTCCGATCTNNN AACTCCCGCCAAGATCAAGAAAG-3′; 120817_LDLR264_r2seq_B, 5′-GGAGTTCAGACGTGTGCTCTTCCGATCTCAG CCTCTTTTCATCCTCCAAGA-3′. For PORCDup20: 120517_PORCN20_r1seq, 5′-CTTTCCCTACACGACGCTCTTCCGATCTNNNC CTCCTACATGG CTTCAGTTTCC-3′; 120517_PORCN20_r2seq, 5′-GGAGTTCAGACGTG TGCTCTTCCGATCTCCAGAGCTCCAAAGAGCAAGTTT-3′. For GLB1Dup84: 120517_GLB184_r1seq, 5′-CTTTCCCTACACGACGCTCTTCCGATCTNNNN AGCCACTCTGGACCTTCTGGTA-3′; 120517_GLB184_r2seq, 5′-GGAG TTCAGACGTGTGCTCTTCCGATCTCCAGTCCGTGAGGATATTGGAAC-3′. For GAADup327/328: 120517_GAA327_r1seq, 5′-CTTTCCCTACACGACGC TCTTCCGATCTNNNGATCGTGAATGAGCTGGTACGTG-3′; 120517_GAA327_r2seq, 5′-GGAGTTCAGACGTGTGCTCTTCCGATCTAACAGCGA GACACAGATGTCCAG-3′.

General HTS data analysis and computational modelling. A detailed and thorough description of methods used for data analysis and computational modelling is available in the Supplementary Methods.

Statistical analysis and reproducibility. Python 2.7 and 3.6 were used to analyse data and perform statistical tests using the SciPy library. Data are represented as mean ± standard error of the mean with 95% confidence intervals. In box plots, box segments show median, 25th and 75th percentiles, whiskers above and below show 1.5 times the interquartile range. Higher and lower points (outliers) are plotted individually or not plotted. Comparison of means of two independent groups was performed using two-sided two-sample t-tests, for which the validity of the normal assumption was analysed using the Shapiro–Wilk tests for small data ($n < 50$ samples) and/or using the Kolmogorov–Smirnov test on larger data ($n > 50$) directly, and/or using the Kolmogorov–Smirnov test on bootstrapped means ($n = 1,000$ bootstrapped samples). In all significance tests performed in the study, the data satisfied our normality criteria for t-tests. For comparison of two independent groups, two-sided two-sample t-tests were used for normally distributed data with equal or similar variance (Student's t-test) or unequal and dissimilar variance (Welch's t-test). A critical value for significance of $P < 0.05$ was used throughout the study.

Here, we report detailed statistical parameters (P value, name of statistical test, test statistic value, degrees of freedom (d.f.), effect size) for all significance tests performed in the study.

Figure 2b, comparison of 1-bp insertion frequencies among Cas9-edited products from 1,996 lib-A target sites. $*P = 5.4 \times 10^{-36}$; $**P = 8.6 \times 10^{-70}$, two-sided two-sample t-test, statistic = −13.0 and −18.4, d.f. = 777 and 1,994; Hedges's $g = 0.94$ and 0.85, for $*$ and $**$, respectively.

Figure 2e, comparison of the 1-bp insertion frequency at sequences in Fig. 2c with varying positions −4 and −3. Box plot as in Figure 2b. $*P = 0.03$; $**P = 2.98 \times 10^{-7}$, two-sided two-sample t-test, statistic = −2.2 and −6.5, d.f. = 185 and 32, Hedges's $g = 0.58$ and 2.3, for $*$ and $**$, respectively.

Figure 3e, comparison of 1-bp insertion frequencies among edited outcomes in U2OS cells ($n = 27$ observations, baseline $n = 1,958$ target sites, $P = 4.2 \times 10^{-8}$, two-sided Welch's t-test, test statistic = 7.56, d.f. = 27.78, Hedges's $g = 1.47$) and HEK293T cells ($n = 26$ observations, baseline $n = 89$ target sites, $P = 8.1 \times 10^{-12}$, two-sided Welch's t-test, test statistic = 10.40, d.f. = 34.14, Hedges' $g = 2.89$).

Extended Data Fig. 3g, box plots displaying total deletion phi score and 1-bp insertion frequencies in mESCs for 312 '4-bp' target sites and 89 VO sequences. $*P = 6.1 \times 10^{-9}$; two-sided two-sample t-test, test statistic = −5.94, d.f. = 399, Hedges' g effect size = 0.49.

Extended Data Fig. 4f, distribution of predicted frameshift frequencies among 1–60-bp deletions for *Sp*Cas9 gRNAs targeting exons ($n = 1,000,294$ gRNAs; mean, 66.4%) and shuffled versions (mean, 69.3%), and introns ($n = 740,759$) in the human genome. Dashed lines indicate means. $***P < 10^{-300}$, two-sided Welch's t-test, test statistic = −145.5, d.f. = 1,506,304, Hedges' $g = −0.19$.

Extended Data Fig. 6a, comparison of microhomology deletions among all deletions at lib-B target sites in wild type ($n = 1,909$ target sites), DPKi3 ($n = 1,999$), MLN4924 ($n = 1,995$), NU7026 ($n = 1,999$), and *Prkdc*$^{-/-}$*Lig4*$^{-/-}$ ($n = 1,446$). Statistical tests performed against wild-type population, Welch's two-sided two-sample t-test. $*P = 5.6 \times 10^{-5}$, test statistic = 4.0, d.f. = 3,870.8, Hedges' g effect size = −0.13. $**P = 3.5 \times 10^{-13}$, test statistic = 7.3, d.f. = 3,890.8, Hedges' g effect size = −0.23. $***P = 5.0 \times 10^{-41}$, test statistic = 13.6, d.f. = 2,651.6, Hedges' g effect size = −0.46.

Extended Data Fig. 6b, comparison of the frequency of each class of microhomology-less deletions among all deletion products in wild-type (lib-A and lib-B target sites, $n = 3,829$ target sites), DPKi3 (lib-B, $n = 1,990$), MLN4924 (lib-B, $n = 1,980$), NU7026 (lib-B, $n = 1,992$) and *Prkdc*$^{-/-}$*Lig4*$^{-/-}$ (lib-A and lib-B target sites, $n = 3,344$). P values are compared to wild type, two-sided Welch's t-test. Comparing among unilateral top strand joining, wild type versus *Prkdc*$^{-/-}$*Lig4*$^{-/-}$ ($P = 1.1 \times 10^{-91}$, test statistic = 20.65, d.f. = 6,223.97, Hedges' $g = 0.50$), versus NU7026 ($P = 4.3 \times 10^{-8}$, test statistic = 5.50, d.f. = 2,798.38, Hedges' $g = 0.18$). Comparing among unilateral bottom strand joining, wild type versus *Prkdc*$^{-/-}$*Lig4*$^{-/-}$ ($P = 4.1 \times 10^{-68}$, test statistic = 17.65, d.f. = 6,479.88, Hedges' $g = 0.42$), versus NU7026 ($P = 7.7 \times 10^{-6}$, test statistic = 4.48, d.f. = 2,868.90, Hedges' $g = 0.50$). Comparing among medial joining, wild type versus MLN4924 ($P = 4.6 \times 10^{-25}$, test statistic = 10.43, d.f. = 3,240.16, Hedges' $g = 0.31$), versus DPKi3 ($P = 4.8 \times 10^{-22}$, test statistic = 9.72, d.f. = 3,231.41, Hedges' $g = 0.29$), versus NU7026 ($P = 4.6 \times 10^{-21}$, test statistic = 9.49, d.f. = 3,130.82, Hedges' $g = 0.29$).

Extended Data Fig. 7f, box plot comparing observed 1-bp insertion frequency in lib-A and 12 pathogenic alleles selected by inDelphi in mESCs (combined data from $n = 2$ independent biological replicates). The box denotes the 25th, 50th and 75th percentiles, whiskers show 1.5 times the interquartile range, and outliers are depicted d.f. = 11.18, Hedges' g effect size = 1.47.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. All data processing, analysis and modelling code is available at www.github.com/gifford-lab/inDelphi-dataprocessinganalysis. The inDelphi model is available online at https://indelphi.giffordlab.mit.edu/.

Data availability

28. Kleinstiver, B. P. et al. Broadening the targeting range of *Staphylococcus aureus* CRISPR–Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).
29. Sherwood, R. I. et al. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014).
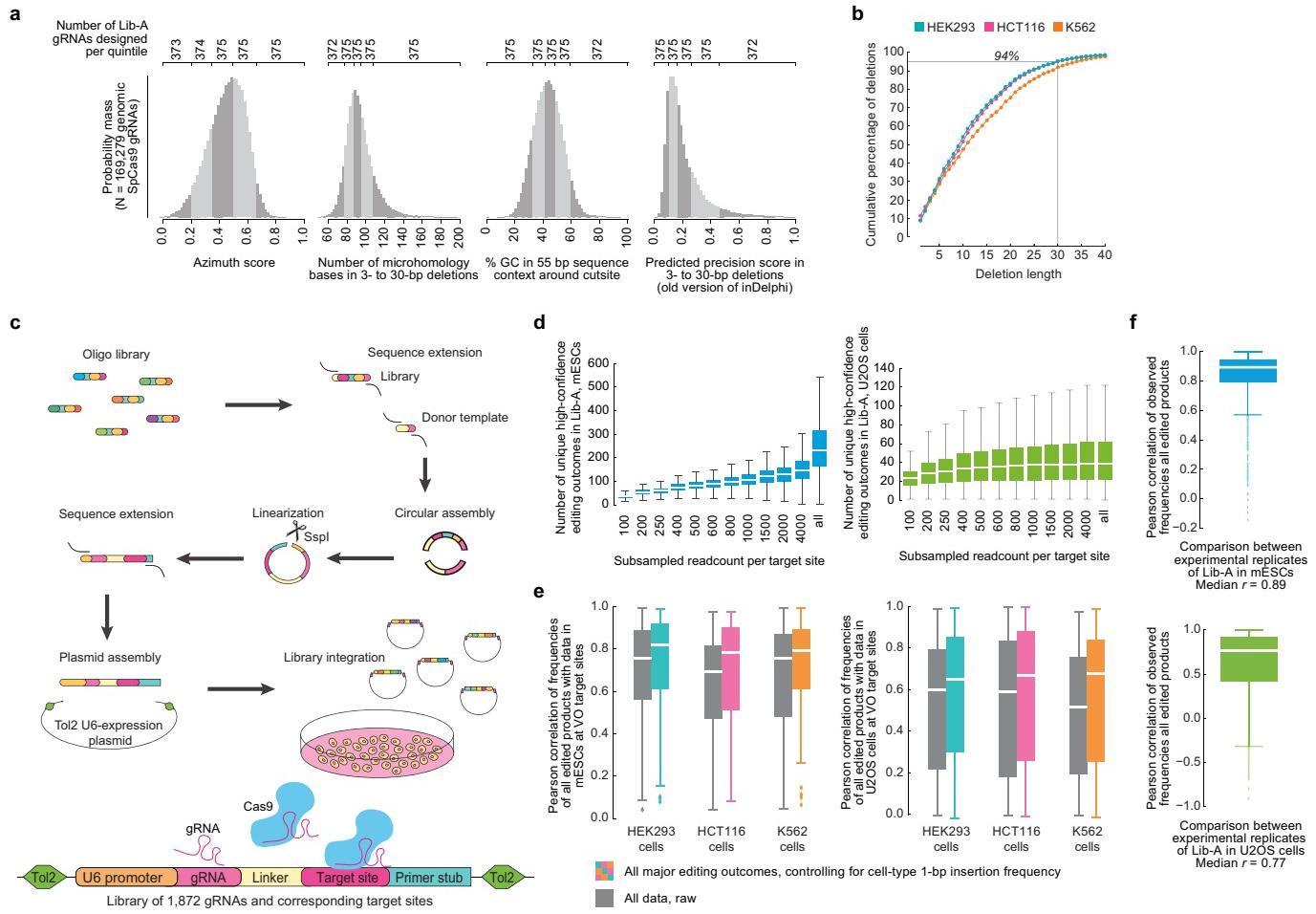
Extended Data Fig. 1 | Design and cloning of a high-throughput library to assess CRISPR–Cas9-mediated editing products, yielding diverse and replicate-consistent data that is concordant with repair spectra at endogenous human genomic loci. a, Empirical distributions of various predicted and measured properties of DNA from 169,279 *Sp*Cas9 gRNA target sites in the human genome. Number of target sites per range used to design lib-A are indicated. **b**, Cumulative percentage of endogenous deletions in VO target sites in HEK293 ($n = 89$ target sites), HCT116 ($n = 92$) and K562 ($n = 86$) cells that delete up to the reported number of nucleotides (x axis). **c**, Schematic of the cloning process used to clone lib-A and lib-B (Methods, Supplementary Discussion, Supplementary Methods). **d**, Number of unique high-confidence editing outcomes (Supplementary Methods) called by simulating data subsampling in data in lib-A ($n = 2,000$ target sites) in mESCs (combined data from $n = 3$ independent

biological replicates) and U2OS cells (combined data from $n = 2$ independent biological replicates). For 'all', the original non-subsampled data are presented. Each box depicts data for 2,000 target sites. Outliers are not depicted. **e**, Pearson's r of genotype frequencies comparing lib-A in mESCs and U2OS cells with endogenous data in HEK293 ($n = 87$ target sites), HCT116 ($n = 88$), and K562 ($n = 86$) cells. Outliers are depicted as diamonds. 1-bp insertion frequency adjustment was performed at each target site by proportionally scaling them to be equal between two cell types. **f**, Pearson's r of genotype frequencies at lib-A target sites, comparing two independent biological replicate experiments in mESCs ($n = 1,861$ target sites, median $r = 0.89$) and U2OS cells ($n = 1,921$, median $r = 0.77$). Outliers are depicted as diamonds. Box plots denote the 25th, 50th and 75th percentiles and whiskers show 1.5 times the interquartile range.
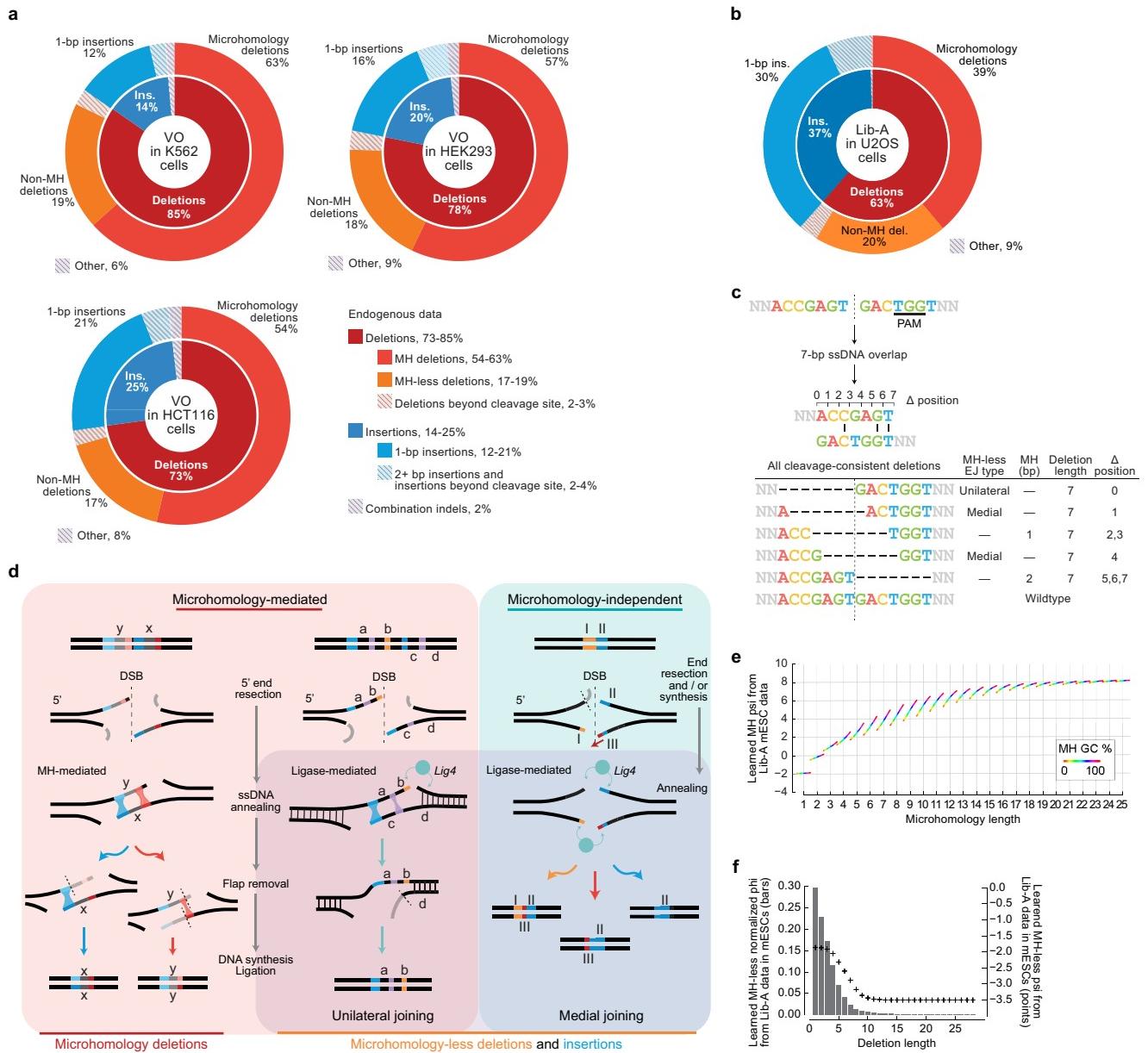
Extended Data Fig. 2 | Categorizing and modelling Cas9-mediated DNA repair products with manual data-analysis and automated machine learning through inDelphi. a, b, Categories of Cas9-mediated genotypic outcomes in data from endogenous contexts at VO target sites in K562 ($n = 88$ target sites), HCT116 ($n = 92$), HEK293 ($n = 89$) cells (collectively, **a**) and U2OS cells (**b**, $n = 1,958$ lib-A target sites). **c,** Categories and defined properties (Supplementary Methods) of all sequence alignments consistent with a Cas9-mediated 7-bp deletion.

d, Hypothesized mechanisms for template-free DNA repair at Cas9-mediated DSBs based on components of the classical NHEJ, alternative NHEJ or MMEJ pathways (Supplementary Discussion). **e,** Function learned for modelling MH deletions (Supplementary Methods). **f,** Function learned for modelling MH-independent deletions (MHless-NN) mapping deletion length to a numeric score (psi, Supplementary Methods, point plot) and with deletion length penalty normalized to sum to 1 (phi, Supplementary Methods, histogram).
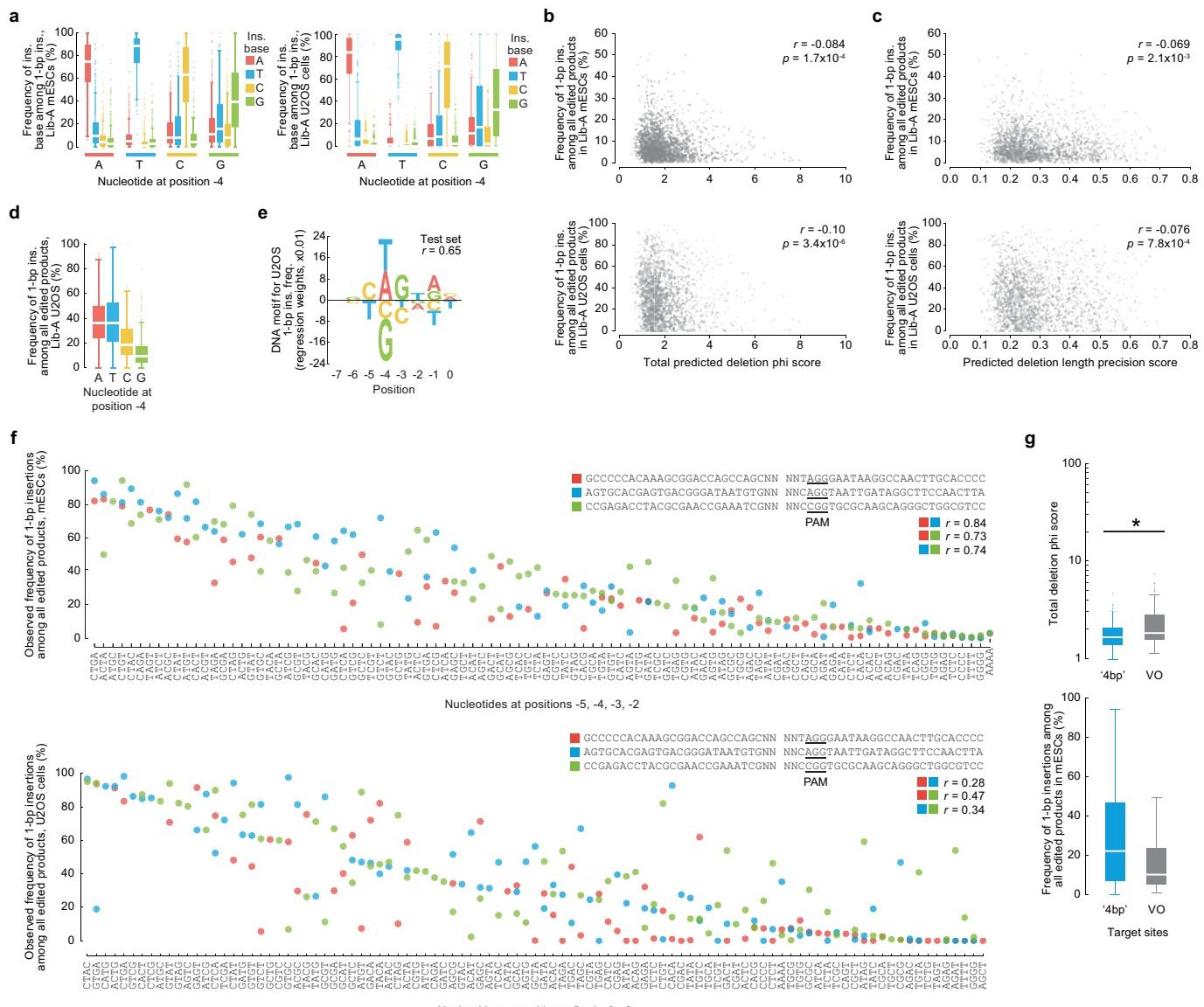
Extended Data Fig. 3 | Influential role of hyperlocal sequence context features in predicting and causing 1-bp insertions. a, Frequency of 1-bp insertions in mESCs ($n = 1,981$ lib-A target sites) and U2OS cells ($n = 1,918$) with varying -4 nucleotides. **b**, **c**, Plot of 1-bp insertion frequency in mESCs ($n = 1,996$ lib-A target sites) and U2OS cells ($n = 1,966$) compared to their total phi score (**b**) and predicted deletion length precision score (**c**) with Pearson's r. **d**, Comparison of 1-bp insertion frequencies among all edited products from 1,966 lib-A target sites in U2OS cells (combined data from $n = 2$ independent biological replicates). **e**, Nucleotides and their effect on the frequency of 1-bp insertions in U2OS cells. Only bases with non-zero linear regression

weights in 10,000-fold iterative cross-validation are shown. Total $n = 1,966$ lib-A target sites. **f**, Insertion frequency in mESCs ($n = 205$) and U2OS cells ($n = 217$) when varying four bases by the cleavage site (positions -5 to -2 counted from the NGG-PAM at positions 0–2) contained within three target sites designed with weak microhomology. **g**, Microhomology strength (deletion phi score) and 1-bp insertions in mESCs for 312 '4-bp' target sites and 89 VO sequences. $*P = 6.1 \times 10^{-9}$; two-sided two-sample t-test, test statistic $= -5.94$, d.f. $= 399$, Hedges' g effect size $= 0.49$. Box plots denote the 25th, 50th and 75th percentiles, whiskers show 1.5 times the interquartile range, and outliers are depicted as diamonds.
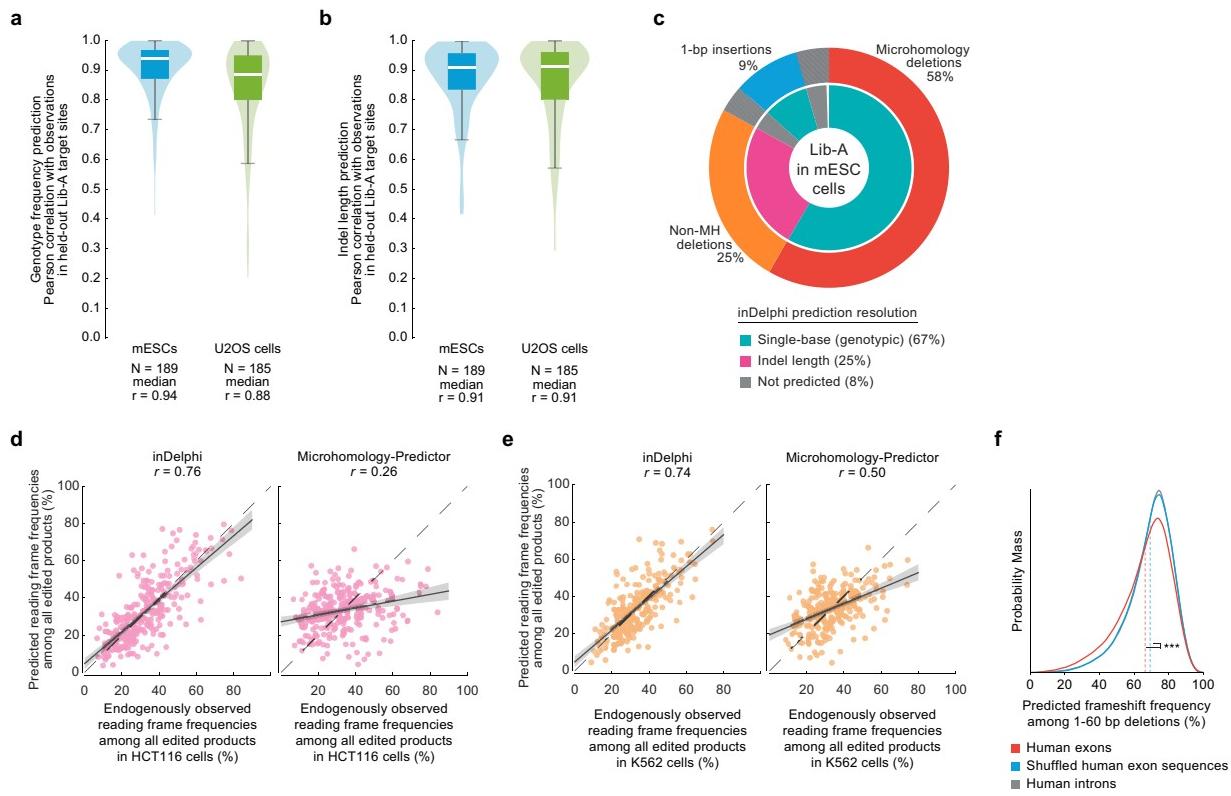
Extended Data Fig. 4 | inDelphi predictions represent nearly all editing outcomes and are accurate at predicting the frequencies of genotypes, indel lengths, and frameshift frequencies. a, b, Pearson's r for held-out lib-A target sites comparing inDelphi predictions with observed frequencies for genotypes (**a**) and indel lengths (**b**) in mESCs and U2OS cells. The box denotes the 25th, 50th and 75th percentiles, whiskers show 1.5 times the interquartile range. Densities were smoothed with noise but do not extend beyond the data. **c,** Pie chart depicting the output of Delphi for specific outcome classes at lib-A target sites in mESCs. **d, e,** Comparison of two methods for frameshift predictions to observed values with Pearson's r in HCT116 cells (**d**, $n = 91$ target sites) and K562 cells (**e**, $n = 82$ target sites). The error band represents the 95% confidence intervals around the regression estimate with 1,000-fold bootstrapping. **f,** Distribution of predicted frameshift frequencies among 1–60-bp deletions for SpCas9 gRNAs targeting exons ($n = 1,000,294$ gRNAs; mean $= 66.4\%$) and shuffled versions (mean, 69.3%), and introns ($n = 740,759$) in the human genome. Dashed lines indicate means. ***$P < 10^{-300}$, two-sided Welch's t-test, test statistic $= -145.5$, d.f. $= 1,506,304$, Hedges' $g = -0.19$.
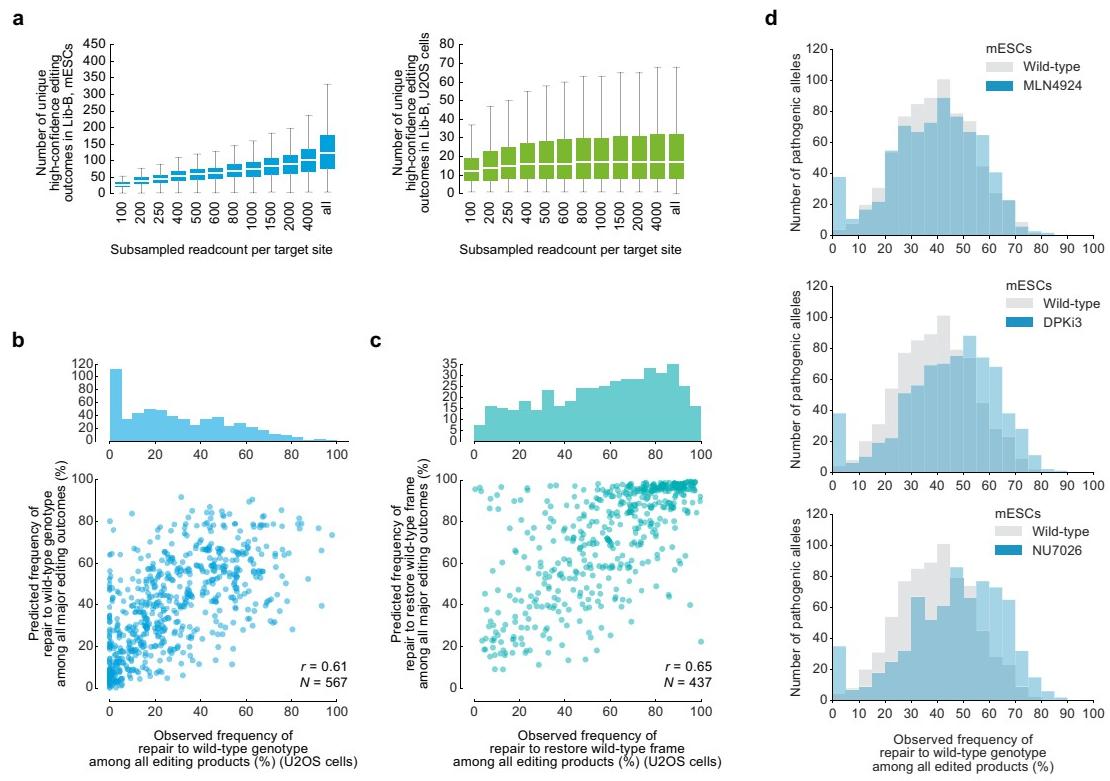
Extended Data Fig. 5 | Characterization of lib-B data including pathogenic microduplication repair in wild-type mESCs, wild-type U2OS cells and mESCs treated with DPKi3, NU7026 and MLN4924. **a**, Box plots of the number of unique high-confidence editing outcomes (see Supplementary Methods) called by simulating data subsampling in data at 2,000 lib-B target sites in mESCs (combined data from $n = 2$ independent technical replicates) and U2OS cells (combined data from $n = 2$ independent biological replicates). In 'all', the full non-subsampled data are presented (see Supplementary Table 2 for read counts). Each box depicts data for 2,000 target sites. The box denotes the 25th, 50th, and 75th percentiles and whiskers show 1.5 times the interquartile range. Outliers are not depicted. **b**, Frequencies of repair to wild-type genotype at 567 ClinVar pathogenic alleles versus predicted frequencies in lib-B in human U2OS cells with Pearson's r. **c**, Frequencies of repair to wild-type frame at 437 ClinVar pathogenic alleles versus predicted frequencies in lib-B in human U2OS cells with Pearson's r. **d**, Frequency of pathogenic microduplication repair in wild-type mESCs ($n = 1,480$ target sites) compared to mESCs treated with MLN4924 ($n = 1,569$), NU7041 ($n = 1,561$) and DPKi3 ($n = 1,563$).
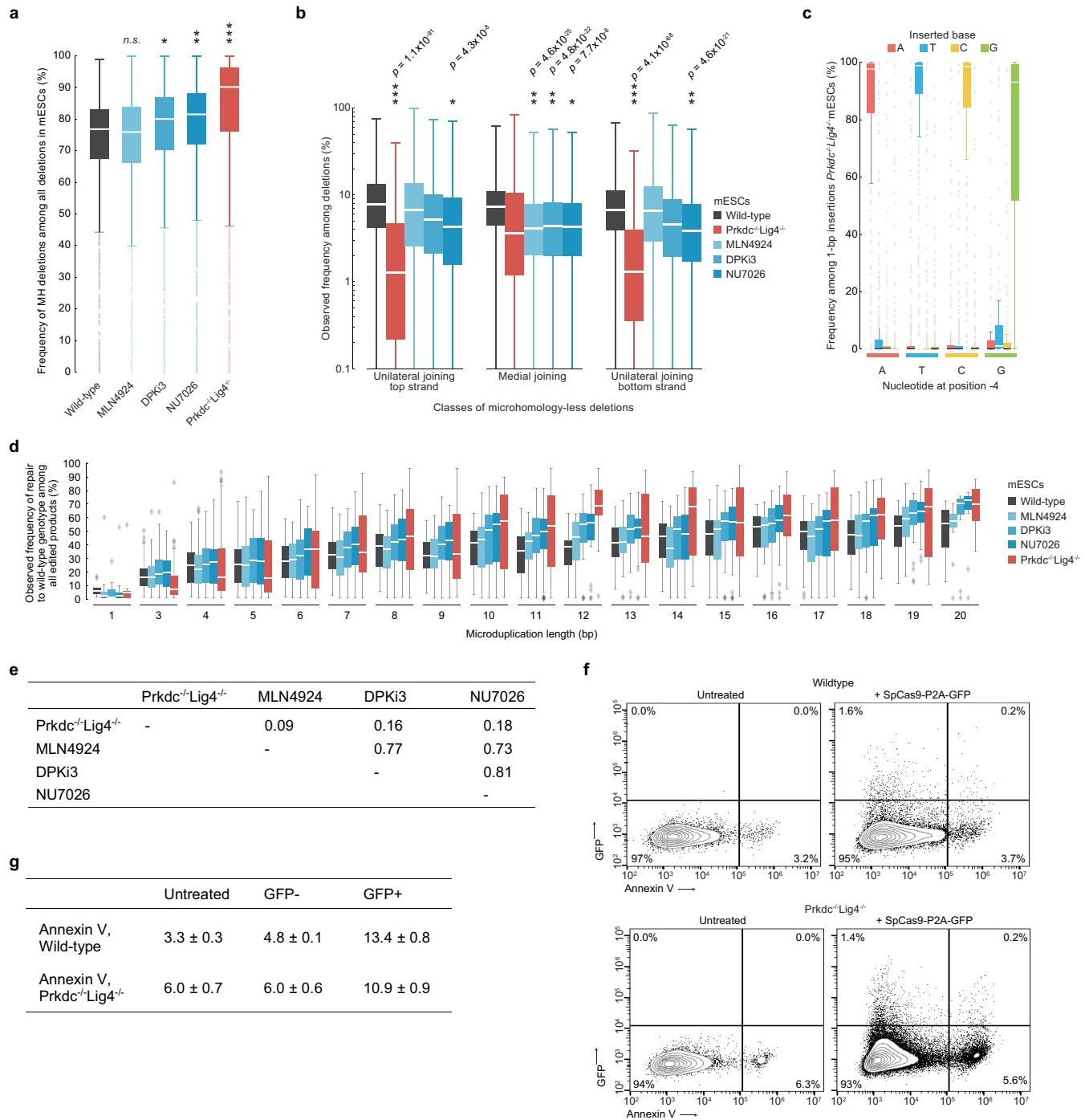
Extended Data Fig. 6 | Altered distributions of Cas9-mediated genotypic products in *Prkdc*[−/−]*Lig4*[−/−] mESCs and mESCs treated with DPKi3, NU7026, and MLN4924 compared to wild-type mESCs. **a**, Comparison of MH deletions among all deletions at lib-B target sites in wild-type cells ($n = 1,909$ target sites), cells treated with DPKi3 ($n = 1,999$), MLN4924 ($n = 1,995$) or NU7026 ($n = 1,999$) and *Prkdc*[−/−]*Lig4*[−/−] cells ($n = 1,446$). Statistical tests performed against wild-type population. $*P = 5.6 \times 10^{-5}$, $**P = 3.5 \times 10^{-13}$, $***P = 5.0 \times 10^{-41}$, two-sided Welch's *t*-test. **b**, Comparison of the frequency of each class of MH-less deletions among all deletion products in wild-type (lib-A and lib-B target sites, $n = 3,829$ target sites), DPKi3 (lib-B, $n = 1,990$), MLN4924 (lib-B, $n = 1,980$), NU7026 (lib-B, $n = 1,992$) and *Prkdc*[−/−]*Lig4*[−/−] (lib-A and lib-B target sites, $n = 3,344$). *P* values are compared to wild-type, two-sided Welch's *t*-test. **c**, Frequency of 1-bp insertions at 1,055 target sites

in lib-A in *Prkdc*[−/−]*Lig4*[−/−] mESCs. **d**, Frequencies of deletion repair to wild-type genotype in lib-B in wild-type mESCs ($n = 1,480$ target sites, combined data from two technical replicates) compared to conditions, with combined data from two independent biological replicates for each of *Prkdc*[−/−]*Lig4*[−/−] ($n = 1,041$ target sites), MLN4924 ($n = 1,569$), NU7026 ($n = 1,561$) and DPKi3 ($n = 1,563$). **e**, Table of Pearson's *r* of the change in disease correction frequency compared to wild-type at $n = 791$ target sites for each pair of conditions. **f**, **g**, Annexin V-568 staining flow cytometry contour plots (**f**) and mean ± standard deviation values (**g**) in wild-type and *Prkdc*[−/−]*Lig4*[−/−] lib-A mESCs following transfection with *Sp*Cas9–P2A–GFP (representative data for $n = 2$ experiments). Box plots denote the 25th, 50th and 75th percentiles, whiskers show 1.5 times the interquartile range, and outliers are depicted as diamonds. For detailed statistics on significance tests, see Methods.

a
HCT116 human CRCs + LDLRwt–P2A–GFP

HCT116 human CRCs + LDLRdup–P2A–GFP
Untreated

+ SaCas9 + gRNA

b
HCT116 human CRCs + LDLRwt–P2A–GFP
HCT116 human CRCs + LDLRdup–P2A–GFP
Untreated

+ SaCas9 + gRNA

c
U2OS + LDLR(526_533dupGGCTCGGA)–P2A–GFP
Untreated

+ SaCas9 + gRNA

U2OS + LDLR(668_681dupAGGACAAATCTGAC)–P2A–GFP
Untreated

+ SaCas9 + gRNA

d
Specimen_002-mES LDLR264 preC

Specimen_002-mES LDLR264 preC

e
Specimen_002-mES LDLR264 CasS

Specimen_002-mES LDLR264 CasS

f

g

Pathogenic 1-bp deletion genotype (hg19)	Observed frequency of 1-bp insertions in all edited products (%)	Observed frequency of repair to wild-type frame in all edited products (%)	Observed frequency of repair to wild-type genotype in all edited products (%)	Observed frequency of repair to wild-type amino acid sequence in all edited products (%)
KLHL7 chr7:23205401	33	55	33	33
NF2 chr22:30038270	17	60	1*	17
NBN chr8:90995063	19	30	17	18
PAX2 chr10:102541067	12	29	0*	11
LDLR chr19:11231158	19	49	13	13
MAN2B1 chr19:12767768	20	50	13	13
MLPH chr2:238436125	22	48	0*	22
SH3BP2 chr4:2824671	26	49	2*	26
COL4A4 chr2:227942769	16	49	16	16
NIPBL chr5:36961683	28	51	0*	28
UGT1A4 chr2:234627639	18	55	17	17
MFRP chr11:119216271	9	42	2*	2

Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Template-free Cas9-nuclease editing of human and mouse cells containing pathogenic alleles. a, **b**, Flow cytometric contour plots showing GFP fluorescence and LDL–Dylight550 uptake in (**a**) and fluorescence microscopy of (**b**) HCT116 cells containing the denoted LDLR alleles and treated with *Sa*Cas9 and gRNA when denoted (representative data for $n = 2$ experiments). **c**, Fluorescence microscopy of U2OS cells containing the denoted LDLR alleles and treated with *Sa*Cas9 and gRNA when denoted (representative data for $n = 2$ experiments). **d**, **e**, Flow cytometry gating strategy used for mESC and LDLRdup–P2A–GFP untreated (**d**) and treated with *Sp*Cas9 and gRNA (**e**). **f**, **g**, Results

of 12 pathogenic 1-bp deletion alleles selected by inDelphi for high 1-bp insertion frequency (combined data from $n = 2$ independent biological replicates) compared to lib-A (**f**) and presented in a table (**g**). The box denotes the 25th, 50th and 75th percentiles, whiskers show 1.5 times the interquartile range, and outliers are depicted as diamonds. $*P = 1.6 \times 10^{-4}$, two-sided Welch's *t*-test. For detailed statistics, see Methods. In the table, the most frequent 1-bp insertion genotype predicted by inDelphi that does not correspond to the wild-type genotype is indicated by an asterisk. In fluorescence microscopy plots, GFP fluorescence is shown in green, LDL–Dylight550 uptake in red, and Hoechst staining nuclei in blue.

Extended Data Table 1 | Frequency of gRNAs in the human genome with denoted Cas9-mediated outcome precision

Precision-X threshold (%)	inDelphi trained on Lib-A data from mESCs for 1-bp ins. module			inDelphi trained on Lib-A data from U2OS cells for 1-bp ins. module		
	Precise product is a deletion (% of gRNAs)	Precise product is a 1-bp insertion (% of gRNAs)	Total % of gRNAs that are precise-X	Precise product is a deletion (% of gRNAs)	Precise product is a 1-bp insertion (% of gRNAs)	Total % of gRNAs that are precise-X
10	82	38	93	70	78	97
15	61	23	75	44	64	87
20	43	15	55	27	53	72
25	30	10	39	17	44	58
30	21	6.6	28	11	36	46
35	15	4.4	19	6.9	28	34
40	10	2.9	13	4.1	21	25
45	6.5	1.9	8.4	2.4	15	18
50	4.3	1.3	5.6	1.4	10	12
55	2.8	0.8	3.6	0.8	6.7	7.5
60	1.8	0.5	2.3	0.5	4.0	4.4
65	1.1	0.3	1.5	0.2	2.2	2.4
70	0.7	0.2	0.9	0.1	1.1	1.2
75	0.4	0.1	0.5	0.04	0.5	0.5
80	0.2	0.08	0.3	0.01	0.2	0.2
85	0.08	0.04	0.1	0.003	0.07	0.08
90	0.03	0.02	0.05	0.0007	0.03	0.03

SpCas9 gRNAs in human exons and introns in mESCs ($n = 1,003,524$ SpCas9 gRNAs) and U2OS cells ($n = 4,498,780$ SpCas9 gRNAs). Predictions were smoothed with Gaussian noise (Supplementary Methods).

Extended Data Table 2 | Endogenous repair of 24 designed high-precision gRNAs in human cell lines

Gene, exon/chr, cutsite (hg19)	Observed frequency among all edited products from deep sequencing at endogenous loci (%)			
	Frameshift, U2OS	Most frequent genotype, U2OS	Frameshift, HEK293T	Most frequent genotype, HEK293T
VEGFA exon1: 458	72, 72	9, 11*	81, 71	28, 9*
VEGFR2 exon5: 2	91, 91	49, 52*	91, 91	49, 23*
PDCD1 exon5: 208	90, 90	20, 22*	91, 91	29, 13*
APOB exon25: 147	83, 83	22, 21*	87, 85	36, 17*
VEGFA exon3: 127	86, 89	28, 30*	92, 91	56, 32*
CCR5 exon1: 1941	83, 81	20, 21*	86, 84	43, 27*
CD274 exon2: 271	85, 86	9, 10*	84, 82	31, 14*
APOB exon26: 5590	91, 89	30, 27*	89	40*
VEGFR2 exon26: 19	82, 82	35, 33*	83, 82	41, 23*
CXCR4 exon1: 825	86, 86	32, 33*	91	55*
PCSK9 exon11: 15	91, 89	64, 64†	89	60†
CCR5 exon1: 885	90, 91	74, 71†	78	65†
CCR5 exon1: 1027	92, 94	62, 62†	91, 92	50, 60†
APOB exon26: 5573	93, 93	75, 74†	93, 95	69, 82†
CCR5 exon1: 61	94, 92	21, 16†	84, 88	19, 28†
CCR5 exon1: 1577	81, 81	29, 30†	80, 84	29, 46†
APOB exon22: 100	89, 90	28, 31†	90, 89	26, 40†
APOBEC3B exon3: 202	83, 83	52, 54†	74, 87	52, 62†
MACCHC chr1: 45973892	97, 95	81, 77†‡	97, 98	79, 86†‡
PROK2 chr3: 71821967	92, 93	45, 45†‡	92, 93	49, 58†‡
IDS chrX: 148564700	96, 95	73, 76†‡	93, 95	63, 79†‡
ECM1 chr1: 150484936	87, 89	47, 52†‡	88, 89	33, 37†‡
KCNH2 chr7: 150644566	46	30†‡	89, 93	71, 75†‡
LDLR chr19: 11222303	91, 92	79, 78†‡	90, 96	78, 84†‡

Data from up to two independent biological replicates are depicted.
*Deletion.
†Insertion.
‡Pathogenic 1-bp insertion allele from Clinvar or HGMD.

Extended Data Table 3 | Repair of ten pathogenic microduplication alleles in individual cellular experiments

Pathogenic allele	LDLRdup1	LDLRdup2	LDLRdup2	LDLRdup3	LDLRdup4	LDLRdup5	PORCNdup	GAAdup	GAAdup	GLB1dup	HPS1dup	ATP7Adup
#AlleleID	245617	245706	245706	245709	245715	246266	25739	354180	354180	98805	ND	ND
Predicted frequency of deletions restoring frame (%)	79	98	96	95	86	94	90	76	93	95	ND	ND
Flow cytometric frameshift frequency (%)	57	95	57	90	72	87	ND	79	74	85	ND	ND
Predicted frequency of repair to wild-type genotype among all major editing products (%)	72	90	83	94	85	86	89	74	91	79	88	43
Flow cytometric phenotypic repair frequency, mESC (%)	36	69	30	53	33	78	ND	ND	ND	ND	ND	ND
Observed frequency of repair to wild-type genotype among all edited products in HTS, mESC (%)	ND	67	39	25	15	65	48	76	59	42	ND	ND
Observed frequency of repair to wild-type genotype among all edited products in HTS, U2OS (%)	100	88	ND	ND	ND	77	ND	ND	ND	ND	ND	ND
Observed frequency of repair to wild-type genotype among all edited products in HTS, HCT116 (%)	ND	ND	ND	24	ND	89	ND	ND	ND	ND	ND	ND
Observed frequency of repair to wild-type genotype among all edited products in Lib-B, mESCs (%)	ND	ND	ND	ND	ND	58	42	ND	63	41	ND	ND
Observed frequency of repair to wild-type genotype among all edited products in primary patient fibroblasts (%)	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	88±14*	98
gRNA sequence	TGCGA AGATG GCTCG GAGGC	GCAAG GACAA ATCTG ACAGG	TCCTC GTCAG ATTTG TCCTG	CTGCA AGGAC AAATC TGAGG	TTTCC TCGTC AGATT TGTCG	ACATC TACTC GCTGG TGAGC	CTGTC CCTGG CTTTT ATCCC	AGCTG CAGAA GGTGA CTGCA	CTGCA GAAGG TGACT GCAGA	TGTGA ACTAT GGTGC ATATA	CAGCA GGGGA GGCCC CCAGC	TTTTT CCATA TAAGA TAAGA
Cas9 Type	KKH SaCas9	KKH SaCas9	KKH SaCas9	KKH SaCas9	KKH SaCas9	SpCas9	SpCas9	KKH SaCas9	SpCas9	SpCas9	SpCas9	SpCas9

ND, not determined. LDLRdup1, LDLR:c.526_533dupGGCTCGGA. LDLRdup2, LDLR:c.668_681dupAGGACAAATCTGAC. LDLRdup3, LDLR:c.669_680dupGGACAAATCTGA. LDLRdup4, LDL-R:c.672_683dupCAAATCTGACGA. LDLRdup5, LDLR:c.1662_1669dupGCTGGTGA. PORCNdup, PORCN:c.1059_1071dupCCTGGCTTTTATC. GAAdup, GAA:c.2704_2716dupCAGAAGGTGACTG. GLB1dup, GLB1:c.1456_1466dupGGTGCATATAT. HPS1dup, HPS:c.1472_1487dupCCTCCCCTGCTGGGGG. ATP7Adup: ATP7A:c.6913_6917dupCTTAT.
*n = 5.

ARTICLE

Structural plasticity of D3–D14 ubiquitin ligase in strigolactone signalling

Nitzan Shabek[1,2,4], Fabrizio Ticchiarelli[3], Haibin Mao[1,2], Thomas R. Hinds[1], Ottoline Leyser[3] & Ning Zheng[1,2]*

The strigolactones, a class of plant hormones, regulate many aspects of plant physiology. In the inhibition of shoot branching, the α/β hydrolase D14—which metabolizes strigolactone—interacts with the F-box protein D3 to ubiquitinate and degrade the transcription repressor D53. Despite the fact that multiple modes of interaction between D14 and strigolactone have recently been determined, how the hydrolase functions with D3 to mediate hormone-dependent D53 ubiquitination remains unknown. Here we show that D3 has a C-terminal α-helix that can switch between two conformational states. The engaged form of this α-helix facilitates the binding of D3 and D14 with a hydrolysed strigolactone intermediate, whereas the dislodged form can recognize unmodified D14 in an open conformation and inhibits its enzymatic activity. The D3 C-terminal α-helix enables D14 to recruit D53 in a strigolactone-dependent manner, which in turn activates the hydrolase. By revealing the structural plasticity of the SCF^{D3-D14} ubiquitin ligase, our results suggest a mechanism by which the E3 coordinates strigolactone signalling and metabolism.

Strigolactones represent a class of plant hormones that regulate a variety of plant growth and developmental processes, such as shoot branching, root development, leaf senescence and flower size[1–3]. Strigolactones are also exuded by plant roots for stimulating interactions with symbiotic fungi[4] and exploited by parasitic plants to time their seed germination[5–8]. As a group of terpenoid lactones, strigolactones typically comprise a butenolide ring (D ring) connected to a variable tricyclic lactone (the ABC rings) via an enol-ether bridge[9,10]. Functional dissection of both natural and synthetic strigolactone molecules has indicated that the C and D rings and their linkage are essential for strigolactone activity, whereas separated ABC or D rings are inactive in plants[10–13].

The perception and signal propagation of strigolactones are coordinated by three highly conserved components: *DWARF3* (*D3*) in rice, or the *Arabidopsis thaliana* orthologue *MAX2* (also known as *AT2G42620*), *D14* (*AT3G03990*) and *D53* (*LOC4349543*) in rice, or the *Arabidopsis* homologues *SMXL6* (*AT1G07200*), *SMXL7* (*AT2G29970*) and *SMXL8* (*AT2G40130*)[12,14–22]. As a member of the α/β serine hydrolase superfamily, D14 not only serves as the strigolactone receptor but also metabolizes strigolactones into tricyclic ABC- and D-ring products, albeit at a rate that is much slower than most known α/β hydrolases[12,13,23]. *D3* in rice (or *MAX2* in *Arabidopsis*) encodes an F-box protein and binds *Arabidopsis* SKP1-like protein (ASK1) to function as a substrate receptor of an SKP1–CUL1–F-box (SCF) ubiquitin ligase complex[20,24]. Recent studies have shown that D3 or MAX2, when bound to D14, mediates the inhibition of shoot branching by sensing strigolactones and ubiquitinating D53 (or the *Arabidopsis* homologues SMXL6, SMXL7 and SMXL8), which is a key nuclear repressor that regulates distinct developmental processes and target genes of strigolactone signalling[15–17,22,25–28].

Early structural studies of strigolactone perception focused on the binding of the hormone to isolated D14 orthologues[12,13,23,29]. Crystal structures of several D14 orthologues—either in their apo or ligand-bound forms—revealed a common α/β fold with a large, solvent-exposed ligand-binding pocket[6,12–14,23,30]. Thus, strigolactones have previously been thought to be perceived by D14 orthologues in this open conformation, although possible conformational changes have

also been suggested. A recent study of the pea (*Pisum sativum*) D14 orthologue RAMOSUS3 (RMS3) suggested that the α/β-fold hydrolase is a single-turnover enzyme, which produces a covalent D-ring-enzyme complex via the catalytic histidine after substrate hydrolysis and the rapid release of the ABC ring[31]. The crystal structure of rice ASK1–D3 in complex with *Arabidopsis* D14 (*At*D14; all uses of D14 without a species prefix refer to D14 from *Oryza sativa*) further uncovered a closed conformation of D3-bound *At*D14, which sequesters the covalently linked intermediate molecule (CLIM) of strigolactone inside a small enclosed pocket[32]. These results raised the possibility that CLIM might represent the active form of the hormone. However, this proposition is complicated by the identification of multiple non-hydrolysable strigolactone agonists[6,33–35].

To better delineate the signalling-competent form of D14 in the context of substrate recognition by the SCF E3 and its relationship with hormone hydrolysis, we have performed structure–function studies of the homogeneous rice D14–D3–D53 system. Our analyses have revealed not only structural plasticity in D3 but also functional states of SCF^{D3-D14} that are switchable by D53 for strigolactone hydrolysis.

Structural plasticity of C-terminal α-helix of D3

We first independently determined the crystal structure of D3 in complex with ASK1. D3 contains an N-terminal F-box motif that forms a canonical interface with ASK1. The C-terminal domain of D3 consists of 20 leucine-rich repeats (LRRs) and adopts a fully circularized solenoid fold; the last LRR (LRR20) of D3 makes direct contact with the three N-terminal LRRs and the C-terminal α-helix (CTH) of ASK1 (Fig. 1a). Distinct from most F-box proteins that contain LRRs, the extreme C-terminal residue of D3 (Asp720) is strictly conserved among diverse plant species (Extended Data Fig. 1a). The backbone and side-chain carboxyl groups of this Asp720 are simultaneously anchored to a positively charged pocket constructed by ASK1, the F-box motif and the LRR domain of D3 (Extended Data Fig. 1b, c).

We subsequently identified a second crystal form of the D3–ASK1 complex, which was crystallized in a different space group with two

[1]Department of Pharmacology, University of Washington, Seattle, WA, USA. [2]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. [3]The Sainsbury Laboratory, University of Cambridge, Cambridge, UK. [4]Present address: Department of Plant Biology, University of California, Davis, Davis, CA, USA. *e-mail: nzheng@uw.edu
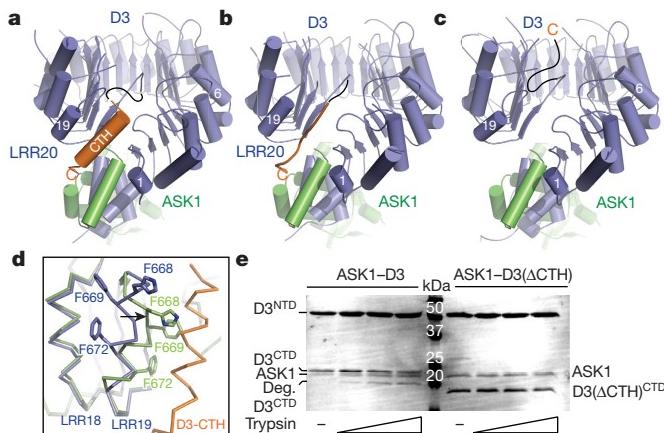
Fig. 1 | Structural plasticity of the D3 C-terminal α-helix. a–c, Overall structures of ASK1 (green) bound to D3 (blue) with its C-terminal α-helix in variable conformation. **d,** Superposition of ASK1–D3 structures shown in **a** (blue) and **c** (green), focusing on LRR18–LRR20. LRR20 (orange) is disordered in the third crystal form (shown in **c**). Black arrow indicates the conformational shift of LRR19 when LRR20 is disordered. **e,** Limited proteolysis of ASK1 in complex with the D3 protein with (ASK1–D3) or without (ASK1–D3(ΔCTH)) the D3-CTH. The experiment was repeated three times. The D3 protein was purified with its N-terminal (NTD) and C-terminal (CTD) segments tightly associated with each other (see Methods). 'Deg.' indicates the proteolytic product of D3-CTD.

Fig. 2 | D3-CTH binds and inhibits D14. a, AlphaScreen assay measuring the ability of D3-CTH peptide to compete with His–D3 for binding the GST–D14 complex. Ten-micromolar GR24 was used to ensure constant binding between D3 and D14 (mean ± s.d. of biological triplicates). **b,** Pull-down assay showing direct GR24-dependent interaction between GST–D3-CTH and His–D14 (experiment repeated three times). WB, western blot. **c,** Biphasic kinetics of YLG hydrolysis by D14. The enzyme was used at a concentration of 0.125 μM to better separate the initial phase from the slow linear phase (purple). An equal amount of extra enzyme was added in a second identical sample (blue) to rule out the possibility that the slow linear phase is due to substrate depletion. **d, e,** Kinetics of YLG hydrolysis by D14 (0.25 μM) ± ASK1–D3 or ± D3-CTH. **f,** Dose–response curve of D3-CTH in inhibiting the enzyme activity of D14.

copies of the complex in the asymmetric unit (Extended Data Table 1). A typical LRR consists of a β-strand, an α-helix and an intervening loop. In contrast to the first D3–ASK1 structure, the LRR20 α-helix in one copy of D3 is mostly reshaped into an extended conformation—although its C-terminal Asp720 residue remains engaged with the basic binding pocket (Fig. 1b, Extended Data Fig. 1d). The electron density of the LRR20 α-helix in the other copy of D3 is entirely missing, which indicates that it is dislodged from the LRR domain and becomes structurally disordered (Fig. 1c). Concurrent with the remodelling of LRR20, the α-helix of the adjacent repeat (LRR19) is shifted away from the LRR18 α-helix towards the space that was originally occupied by the LRR20 α-helix (Fig. 1d). In a similar but non-isomorphous crystal, the electron density of the LRR20 α-helix is absent in both copies of the complex (Extended Data Table 1). Together, these structural results strongly suggest that the CTH of D3 (D3-CTH) can have a dynamic topology that is capable of switching between engaged and dislodged states. Using limited proteolytic digestion, we further verified that the C-terminal end of D3 is more sensitive to protease cleavage than the rest of the protein (Fig. 1e, Extended Data Figs. 2, 3). Therefore, the conformational plasticity of the CTH of D3 is an inherent property of the F-box protein in solution.

D3–CTH binds and inhibits D14

This unusual structural feature of D3-CTH prompted us to investigate its role in the D3–D14 interaction. We first established a quantitative method for measuring the GR24-dependent interaction between D3 and D14 in an AlphaScreen assay (Extended Data Fig. 4a). In a dose-dependent manner, a 28-amino-acid peptide of D3-CTH was able to compete with D3 for binding D14 at a saturating concentration of GR24 (Fig. 2a). When fused to glutathione S-transferase (GST), D3-CTH robustly pulled down D14 in a GR24-dependent manner (Fig. 2b), indicating that it can directly interact with the α/β hydrolase. In the previously reported crystal structure of the AtD14–D3–ASK1 complex, D3-CTH is fully engaged with the LRR domain[32]. D3-CTH uses its N-terminal tip and preceding loop to assist the recognition of the CLIM-bound α/β hydrolase by three other D3 C-terminal LRRs (LRR17–LRR19) (Extended Data Fig. 4b). Without the rest of the F-box protein, it is unlikely that our D3-CTH peptide interacts with CLIM-bound D14 in a similar manner.

D14 hydrolyses the fluorogenic strigolactone agonist Yoshimulactone Green (YLG)[7] with a biphasic time course that is characterized by a rapid initial phase followed by slow linear hydrolysis (Fig. 2c). Such a two-stage reaction has been reported for RMS3, which becomes completely inhibited by CLIM after hydrolysing a substrate molecule[31]. Instead of being a single-turnover enzyme, however, D14 slowly released the D-ring under our experimental conditions and continued to hydrolyse additional substrate, as evidenced by the slow linear phase of its enzyme kinetics. Consistent with the recognition and stabilization of CLIM-bound D14 by D3–ASK1, the addition of recombinant D3–ASK1 to D14 reduced the substrate hydrolysis rate of D14 in the slow linear phase without compromising the rapid initial reaction (Fig. 2d). An increasing amount of the isolated D3 C-terminal peptide not only blocked the slow linear hydrolysis but also inhibited the initial reaction—that is, the first cycle of YLG hydrolysis by D14 (Fig. 2e). Moreover, the half maximal inhibitory concentration of the D3 C-terminal peptide in inhibiting D14 and its affinity to the α/β hydrolase are in the same range (Fig. 2a, f). These results suggest that the C-terminal region of D3—when dislodged from the LRR domain—can interact with and block the enzymatic activity of D14 in a manner that is different from the engaged form of this C-terminal region.

Structure of a D3-CTH–D14–GR24 complex

To map the binding mode of a D3 C-terminal peptide to D14 bound to GR24, we crystallized and determined the structure of D14 that is N-terminally fused with a D3 C-terminal sequence in the presence of GR24. In the crystal, the D3 C-terminal sequence in one polypeptide chain acts *in trans* and interacts with D14 of a neighbouring molecule, which contains an island of electron density near the catalytic triad of D14 (Extended Data Fig. 5a). Similar to the isolated form of D14, D14 bound to D3-CTH adopts an open conformation that is characterized by a solvent-accessible substrate-binding pocket (Fig. 3a). The electron density inside the pocket matches the overall shape of the GR24 D ring, although the position of this density is different from where the GR24 degradation product—5-hydroxy-3-methylbutenolide (D-OH)—has previously been found[13,23] (Extended Data Fig. 5b, c). We tentatively assigned this density to the D ring of an unhydrolysed hormone for several reasons. First, the crystallized D14 fusion polypeptide shows little enzymatic activity in hydrolysing substrate (Extended Data Fig. 5d).
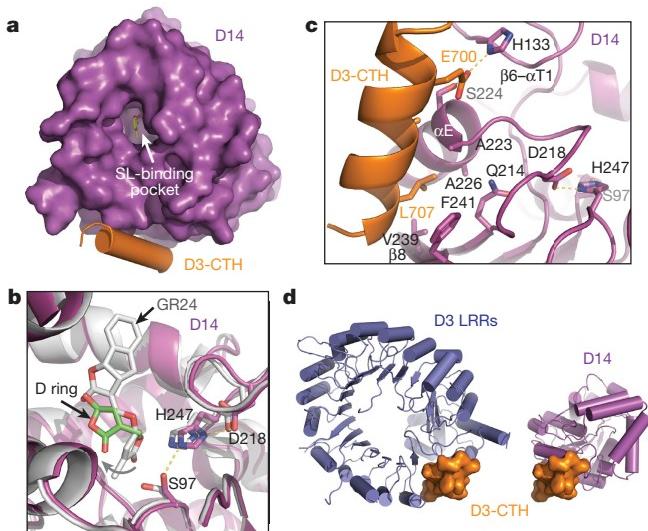
Fig. 3 | Structure of the D3-CTH–D14–GR24 complex. a, Top view of D14 (purple) bound to D3-CTH (orange), with an open ligand-binding pocket. SL, strigolactone. **b**, Close-up view of the ligand-binding pocket of D14 bound to D3-CTH (magenta), superimposed with D14 bound to GR24 (grey, RCSB Protein Data Bank code (PDB) 5DJ5). The catalytic triad of D14 is shown in sticks. A curved arrow indicates the removal of the GR24 D ring away from the catalytic serine residue in the D14 bound to D3-CTH. **c**, Interface between D3-CTH (orange) and D14 (magenta) with interacting residues shown in sticks. **d**, A comparison of the structural context of D3-CTH (orange) in the ASK1–D3 complex (crystal form 1; blue, left) and the D14–D3-CTH complex (magenta, right).

Second, the electron density is extended beyond the D ring, and points towards the exit of the hormone-binding pocket. Third, the location of the D ring predicts that the tricyclic ABC rings are largely solvent-exposed, which could explain their missing electron density. In comparison to the previously reported GR24–D14 structure[14], the hormone is markedly removed from the catalytic centre instead of being poised for hydrolysis (Fig. 3b). The relative position of the hormone, and its orientation to the active site, suggest that it is bound to the enzyme in a non-reactive configuration.

Upon binding to D14, the D3-CTH sequence adopts the same α-helical conformation as seen in its engaged form (Fig. 3a, d). D3-CTH docks to a surface site on D14 that is opposite to where D3 binds in the structure of the *At*D14–D3–ASK1 complex (Extended Data Fig. 5e). At one end of this interface, Glu700 clamps the D3 α-helix to the hydrolase by making polar interactions with Ser224 and His133 on the D14 αE helix and β6–αT1 loop, respectively. At the other end of the interface, D3-CTH inserts Leu707 into a hydrophobic cleft formed between the αE helix and β8 strand of D14 (Fig. 3c). As a whole, the helical portion of the D3 C-terminal sequence buries a total of 800 Å² surface area on D14. If the acidic C terminus of D3-CTH were not fused to the N terminus of the neighbouring D14 molecule, it might be able to interact with a nearby basic D14 surface (Extended Data Fig. 5f).

Superposition analysis of free D14 and D14 bound to D3-CTH reveals a slight rotation of the cap domain around the hormone-binding pocket, which could couple the docking of D3-CTH to the binding of the unhydrolysed hormone (Extended Data Fig. 5g). A closer comparison of all D14 structures also reveals a potential allosteric pathway that links D3-CTH binding to the D14 catalytic triad (Extended Data Fig. 5h). Importantly, D3-CTH uses several common residues to either bind D14 in its open conformation or engage with the rest of the LRR domain (Fig. 3d). The incompatibility of the two structures strongly suggests that D3-CTH binds D14 when dislodged from the LRR domain. Overall, the binding mode of the D3 C-terminal peptide to D14 reflects a functional state of SCF$^{D3–D14}$ that is different from D14 bound to CLIM.

Reactivation of D3–bound D14 by D53

In an in vitro protein degradation system, we next reconstituted proteasome-mediated degradation of recombinant D53 with cell-free extracts prepared from *Arabidopsis* Col-0 seedlings (Fig. 4a). Consistent with the essential role of the MAX2 in strigolactone signalling, *max2-1* extracts lack D53-degrading activity but can be rescued by the addition of recombinant D3 and D14. On the basis of its sequence homology with proteins of the class I Clp ATPase family, D53 is predicted to contain an N-terminal domain and two putative ATPase domains (D1 and D2). We purified each of these D53 domains fused to GST and found that the D2 domain of D53 (D53-D2) is solely responsible for binding D14 in a GR24-dependent manner[16] (Extended Data Fig. 6a). Both full-length D53 and the isolated D2 domain can form a stable complex with D14–D3–ASK1 in the presence of GR24 as detected by size-exclusion chromatography (Fig. 4b, Extended Data Fig. 6b). Although previous studies have suggested that D14 and D3 can individually interact with D53[15–17], we found that the D2 domain of D53 becomes stably associated with D14 only in the presence of D3 and GR24. The three binding partners, therefore, assemble cooperatively into a ternary complex, which explains the degradation of the D2 domain of recombinant D53 by the proteasome in a *MAX2*-dependent manner (Fig. 4a). Together, these data pinpoint the D2 domain of D53 as the functional module for hormone-induced and SCF$^{MAX2–D14}$-catalysed turnover.

We next used the D2 domain of D53 to probe the role of the C-terminal region of D3 in D14-mediated substrate binding. Consistent with the ability of D3 to flip out its CTH without compromising its structural integrity, truncating the 28-amino-acid C-terminal region had no detectable effect on the folding and solution behaviour of D3 (Extended Data Fig. 6c). However, the C-terminally truncated D3 mutant protein could neither form a ternary complex with D14 and D53-D2 on a sizing column, nor restore the D53 degradation activity of the *max2-1* extracts (Fig. 4a, Extended Data Fig. 4c, d), which indicates a critical role of the C-terminal region of D53 in substrate recruitment by SCF$^{D3–D14}$. The isolated C-terminal peptide of D3 was able to stimulate D14 and D53-D2 to pull down one another in the presence of GR24 (Fig. 4c, Extended Data Fig. 7a). In the more-quantitative AlphaScreen assay, the D3 peptide—but not two shorter versions—elicited the same effect in a dose-dependent manner (Extended Data Fig. 7b). Mutation of a single D14 residue (S224E) at the interface revealed in our D3-CTH–D14 structure compromised D3-CTH–D14 binding and was sufficient to prevent GST–D53-D2 from pulling down D3 (Extended Data Fig. 7c, d). These data strongly suggest that the C-terminal region of D3 helps recruit D53 when D3-CTH is liberated from the LRR domain of D3 and becomes compatible for binding D14 with the canonical open conformation. This notion is further corroborated by the impaired D53 degradation observed with either the isolated D3-CTH peptide or the D14(S224E) mutation in the cell-free extracts (Extended Data Fig. 7e, f). *D53* was originally identified through the gain-of-function rice mutant *d53*, the gene product of which becomes resistant to strigolactone-induced degradation owing to the loss of four amino acids in the D2 domain[15,16]. Accordingly, the D2 domain of the recombinant d53 mutant protein was unable to pull down D14 in the presence of D3-CTH and GR24, and remained stable in Col-0 cell extracts (Extended Data Fig. 7a, g). These results further support the functional relevance of the D3-CTH-mediated recruitment of D53 to SCF$^{D3–D14}$.

Given the structural flexibility of D14, we next investigated the effect of substrate binding on the hydrolase activity of D14. Similar to GR24, YLG can induce complex assembly among D14, D3–ASK1 and the D2 domain of D53 (Extended Data Fig. 7h). By monitoring YLG hydrolysis, we detected little change in the enzymatic kinetics of D3–ASK1-arrested D14 when the D2 domain of D53 was present (Extended Data Fig. 7i). By contrast, the addition of D53 robustly blocked the inhibition of the enzymatic activity of D14 by the D3 C-terminal peptide, both in the rapid initial reaction and the slow linear phase (Fig. 4d). Together, these results suggest that the enzymatic activity of D14 within the SCF$^{D3–D14}$ ubiquitin ligase complex is susceptible to modulation
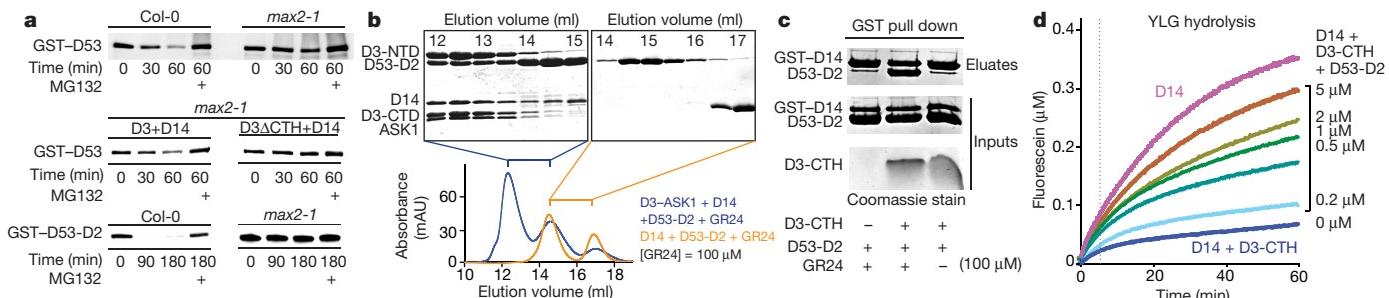
Fig. 4 | Interactions among D3-CTH, D53 and D14. a, Time-dependent degradation of GST–D53 and GST–D53-D2 in *Arabidopsis* seedlings of Col-0 and *max2-1* mutant extracts, with and without recombinant D14 and either D3 or D3(ΔCTH). MG132, proteasome inhibitor. **b**, Size-exclusion chromatography analysis of the interaction between the D2 domain of D53 and D14–GR24 ± ASK1–D3, with sodium dodecyl sulfate–polyacrylamide gel electrophoresis analysis of the elution fractions. **c**, GST pull-down assay using recombinant GST–D14 and non-tagged D53-D2 ± D3-CTH and/or GR24. **d**, Kinetics of YLG hydrolysis by D14 in the presence of D3-CTH at increasing concentrations. All experiments repeated three times.

by D53 binding, and such a modulatory effect is dependent on the conformational state of the CTH of the F-box protein.

SMXL7 levels are compromised by D3–CTH

To further validate the role of D3-CTH in vivo, we expressed SMXL7–YFP alone or in combination with *At*D14, *At*D14 fused to MAX2-CTH (*At*D14-CTH), or MAX2-CTH (CTH) alone in tobacco epidermal cells. Despite the cross-species reactions, SMXL7 was markedly destabilized upon GR24 treatment (23% reduction, Fig. 5a, e), which indicates that the strigolactone perception machinery that is endogenous to tobacco epidermal cells is sufficient to induce GR24-dependent degradation of SMXL7. In support of the functionality of *At*D14 in tobacco epidermal cells, the above response was further accentuated in nuclei that co-express *At*D14-CTH–mCherry—reaching a nearly 50% reduction in the level of SMXL7 by the end of the incubation time (Fig. 5b, e). However, this enhancement effect was completely eliminated and reversed to 11.8% and 6% reduction when SMXL7–YFP was co-expressed with D14-CTH–mCherry or CTH-NLS (nuclear localization sequence)–mCherry, respectively (Fig. 5c–e). The MAX2-CTH, therefore, not only prevented *At*D14 from accelerating the degradation of

SMXL7 but also impaired the destabilization of SMXL7 by endogenous strigolactone-signalling components, in response to GR24 treatment.

A model of the functional states of SCF^{D3-D14}

Our studies have uncovered a structural plasticity in the D3 F-box protein, which can adopt two distinct structural states by altering the topology of its CTH. With an engaged CTH, the F-box protein is structurally compatible for binding the inactive closed conformation of CLIM-bound D14. When the D3-CTH is unleashed from the LRR domain, D3 uses the helical structural element to capture hormone-bound D14 via a different interface. In this binding mode, the hydrolase maintains its open conformation, which allows its enzymatic activity to be tunable by D53, the substrate of the SCF E3. We postulate that the plant SCF^{D3-D14} complex has evolved these unusual features to orchestrate strigolactone sensing, substrate polyubiquitination and hormone metabolism in a highly coordinated manner. To explain the activity of non-hydrolysable strigolactone agonists[6,33–35], we propose that D14 perceives and transduces the hormonal signal in its open conformation, which is recognized by the D3-CTH and is competent for D53 binding (Extended Data Fig. 8, Supplementary Discussion). Before loading the SCF substrate, D3 arrests strigolactone-bound D14 to prevent premature hydrolysis of the hormone. Hormone-dependent association of D53 with SCF^{D3-D14} not only triggers D53 polyubiquitination but also licenses D14 to catalyse strigolactone metabolism, which takes place while D53 is being fully modified (or after D53 has been fully modified) by a ubiquitin chain. A more-detailed quantitative understanding of the timing of strigolactone hydrolysis and polyubiquitin chain assembly on a substrate awaits further studies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0743-5.
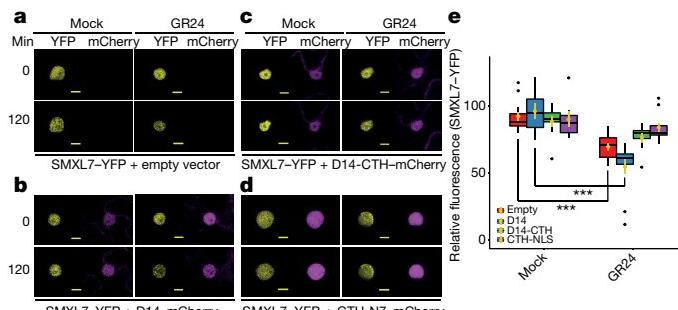
Fig. 5 | SMXL7–YFP stability in response to GR24 and D3-CTH expression. a–d, Representative images of levels of SMXL7 in response to GR24 application in tobacco epidermal cells. Levels of SMXL7–YFP (yellow) at 0 and 120 min post-treatment are shown for single nuclei that co-express empty vector (**a**), D14–mCherry (**b**), D14-CTH–mCherry (**c**) or CTH-NLS–mCherry (**d**), all of which are displayed in magenta. Scale bars, 10 μm. **e**, Relative SMXL7 abundance at 2 h post-treatment, comparing GR24 and mock treatment in cells that either express SMXL7–YFP alone or co-express SMXL7 and D14–mCherry, D14-CTH–mCherry or CTH-NLS–mCherry, respectively. Yellow dots and bars are mean ± s.e.m. ***$P < 0.001$, $n = 7$ (nuclei), two-tailed Student's t-test. Black dots, data more than 3 s.d. from mean. Coloured boxes represent the central 50% of the distribution, with the median shown as a horizontal bar. Top and bottom vertical bars represent 75–100% and 0–25% of the distribution of data points, respectively. P values for empty vector, D14, D14-CTH and CTH-N7 (N7 is a nuclear localization sequence) are 1.294×10^{-5}, 7.188×10^{-4}, 4.401×10^{-2} and 3.86×10^{-1}, respectively.

1. Al-Babili, S. & Bouwmeester, H. J. Strigolactones, a novel carotenoid-derived plant hormone. *Annu. Rev. Plant Biol.* **66**, 161–186 (2015).
2. Waters, M. T., Gutjahr, C., Bennett, T. & Nelson, D. C. Strigolactone signaling and evolution. *Annu. Rev. Plant Biol.* **68**, 291–322 (2017).
3. Gomez-Roldan, V. et al. Strigolactone inhibition of shoot branching. *Nature* **455**, 189–194 (2008).
4. Akiyama, K., Matsuzaki, K. & Hayashi, H. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **435**, 824–827 (2005).
5. Cook, C. E., Whichard, L. P., Turner, B., Wall, M. E. & Egley, G. H. Germination of witchweed (*Striga lutea* lour.): isolation and properties of a potent stimulant. *Science* **154**, 1189–1190 (1966).
6. Toh, S. et al. Structure–function analysis identifies highly sensitive strigolactone receptors in *Striga*. *Science* **350**, 203–207 (2015).
7. Tsuchiya, Y. et al. Probing strigolactone receptors in *Striga hermonthica* with fluorescence. *Science* **349**, 864–868 (2015).

8. Conn, C. E. et al. Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* **349**, 540–543 (2015).
9. Xie, X. et al. Confirming stereochemical structures of strigolactones produced by rice and tobacco. *Mol. Plant* **6**, 153–163 (2013).
10. Zwanenburg, B., Mwakaboko, A. S., Reizelman, A., Anilkumar, G. & Sethumadhavan, D. Structure and function of natural and synthetic signalling molecules in parasitic weed germination. *Pest Manag. Sci.* **65**, 478–491 (2009).
11. Akiyama, K., Ogasawara, S., Ito, S. & Hayashi, H. Structural requirements of strigolactones for hyphal branching in AM fungi. *Plant Cell Physiol.* **51**, 1104–1117 (2010).
12. Hamiaux, C. et al. DAD2 is an α/β hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. *Curr. Biol.* **22**, 2032–2036 (2012).
13. Nakamura, H. et al. Molecular mechanism of strigolactone perception by DWARF14. *Nat. Commun.* **4**, 2613 (2013).
14. Zhao, L. H. et al. Destabilization of strigolactone receptor DWARF14 by binding of ligand and E3-ligase signaling effector DWARF3. *Cell Res.* **25**, 1219–1236 (2015).
15. Jiang, L. et al. DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature* **504**, 401–405 (2013).
16. Zhou, F. et al. D14-SCF^{D3}-dependent degradation of D53 regulates strigolactone signalling. *Nature* **504**, 406–410 (2013).
17. Wang, L. et al. Strigolactone signaling in *Arabidopsis* regulates shoot development by targeting D53-like SMXL repressor proteins for ubiquitination and degradation. *Plant Cell* **27**, 3128–3142 (2015).
18. Nelson, D. C. et al. F-box protein MAX2 has dual roles in karrikin and strigolactone signaling in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **108**, 8897–8902 (2011).
19. Arite, T. et al. *d14*, a strigolactone-insensitive mutant of rice, shows an accelerated outgrowth of tillers. *Plant Cell Physiol.* **50**, 1416–1424 (2009).
20. Stirnberg, P., Furner, I. J. & Ottoline Leyser, H. M. MAX2 participates in an SCF complex which acts locally at the node to suppress shoot branching. *Plant J.* **50**, 80–94 (2007).
21. Beveridge, C. A. & Kyozuka, J. New genes in the strigolactone-related shoot branching pathway. *Curr. Opin. Plant Biol.* **13**, 34–39 (2010).
22. Soundappan, I. et al. SMAX1-LIKE/D53 family members enable distinct MAX2-dependent responses to strigolactones and karrikins in *Arabidopsis*. *Plant Cell* **27**, 3143–3159 (2015).
23. Zhao, L. H. et al. Crystal structures of two phytohormone signal-transducing α/β hydrolases: karrikin-signaling KAI2 and strigolactone-signaling DWARF14. *Cell Res.* **23**, 436–439 (2013).
24. Zheng, N. & Shabek, N. Ubiquitin ligases: structure, function, and regulation. *Annu. Rev. Biochem.* **86**, 129–157 (2017).
25. Liang, Y., Ward, S., Li, P., Bennett, T. & Leyser, O. SMAX1-LIKE7 signals from the nucleus to regulate shoot development in *Arabidopsis* via partially EAR motif-independent mechanisms. *Plant Cell* **28**, 1581–1601 (2016).
26. Song, X. et al. IPA1 functions as a downstream transcription factor repressed by D53 in strigolactone signaling in rice. *Cell Res.* **27**, 1128–1141 (2017).
27. Stanga, J. P., Smith, S. M., Briggs, W. R. & Nelson, D. C. *SUPPRESSOR OF MORE AXILLARY GROWTH2 1* controls seed germination and seedling development in *Arabidopsis*. *Plant Physiol.* **163**, 318–330 (2013).
28. Ma, H. et al. A D53 repression motif induces oligomerization of TOPLESS corepressors and promotes assembly of a corepressor–nucleosome complex. *Sci. Adv.* **3**, e1601217 (2017).
29. Abe, S. et al. Carlactone is converted to carlactonoic acid by MAX1 in *Arabidopsis* and its methyl ester can directly interact with AtD14 *in vitro*. *Proc. Natl Acad. Sci. USA* **111**, 18084–18089 (2014).
30. Kagiyama, M. et al. Structures of D14 and D14L in the strigolactone and karrikin signaling pathways. *Genes Cells* **18**, 147–160 (2013).
31. de Saint Germain, A. et al. An histidine covalent receptor and butenolide complex mediates strigolactone perception. *Nat. Chem. Biol.* **12**, 787–794 (2016).
32. Yao, R. et al. DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature* **536**, 469–473 (2016).
33. Toh, S., Holbrook-Smith, D., Stokes, M. E., Tsuchiya, Y. & McCourt, P. Detection of parasitic plant suicide germination compounds using a high-throughput *Arabidopsis* HTL/KAI2 strigolactone perception system. *Chem. Biol.* **21**, 988–998 (2014).
34. Lumba, S., Subha, A. & McCourt, P. Found in translation: applying lessons from model systems to strigolactone signaling in parasitic plants. *Trends Biochem. Sci.* **42**, 556–565 (2017).
35. Lumba, S., Bunsick, M. & McCourt, P. Chemical genetics and strigolactone perception. *F1000Res.* **6**, 975 (2017).

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0743-5.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0743-5.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to N.Z.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Protein preparation and purification. The full-length rice D3 (*O. sativa*) and *A. thaliana* ASK1 were co-expressed as a $6 \times$ His–$2 \times$ Msb (msyB)[36] fusion protein and an untagged protein, respectively, in Hi5 suspension insect cells. The ASK1–D3 complex was isolated from the soluble cell lysate by Q Sepharose High Performance resin (GE healthcare). NaCl eluates (500 mM) were subjected to Nickel Sepharose Fast Flow (GE Healthcare) and were eluted with 250 mM imidazole. To remove the $6 \times$ His–$2 \times$ Msb fusion tag, the clarified complex was cleaved at 4 °C for 16 h by tobacco etch virus (TEV) protease, and was further purified by anion exchange and gel-filtration chromatography. For crystallization and biochemical analysis purposes, the D3-expressing construct was designed to eliminate a non-conserved 40-residue disordered loop between amino acid 476 and amino acid 514 after affinity purification. The resulting D3 fusion protein contains an $6 \times$ His–$2 \times$ Msb tag at the N terminus and three TEV protease sites: between the Msb tag and D3, after T476 and before L514, yielding a purified split form of D3 with D3 N-terminal domain (1–476) and C-terminal domain (514–720) stably associated (Extended Data Figs. 2, 3). D3(ΔCTH) (*O. sativa*, residues 1–693) was co-expressed with ASK1 and purified in the same manner as full-length D3. Purified ASK1–D3 and ASK1–D3(ΔCTH) complexes were independently eluted as a single mono-dispersive peak off a Superdex-200 gel-filtration column (GE Healthcare) with an estimated molecular weight of 93 kDa or 90 kDa, respectively. The D3(ΔC10) construct, which lacks the C-terminal 10 amino acids, was also purified using a similar procedure. Rice D14 protein (*O. sativa*, residues 52–318) was expressed as a $6 \times$ His–SUMO fusion protein from the expression vector pSUMO (LifeSensors, and a gift from E. Xu). BL21 (DE3) cells transformed with the expression plasmid were grown in LB broth at 16 °C to an OD_{600} of ~1.0 and induced with 0.1 mM IPTG for 16 h. Cells were collected, re-suspended and lysed in extract buffer (20 mM Tris–HCl, pH 8.0, 200 mM NaCl). His–SUMO–D14 was isolated from soluble cell lysate by Ni–NTA resin. The eluted His–SUMO–D14 was subjected to anion exchange and the eluted His–SUMO–D14 was cleaved overnight with SUMO protease (Ulp1, LifeSensors) at a protease-to-protein ratio of 1:1,000 at 4 °C. The cleaved His–SUMO tag was removed by passing through a Nickel Sepharose column, and the protein was further purified by chromatography through a Superdex-200 gel-filtration column in 20 mM Tris, pH 8.0, 200 mM NaCl, 2 mM DTT. Full-length D53 (*O. sativa*) was expressed as a GST fusion protein in Hi5 suspension insect cells. D14 (*O. sativa*, residues 52–318), D3-CTH (*O. sativa*, residues 693–720), D53 N domain (D53-N, residues 1–181), D53 D1 domain (D53-D1, residues 182–406), D53 D2 domain (D53-D2, residues 718–1,131), and the D2 domain of the d53 mutant (F811T followed by deletion of residues 812–818, as previously described[15,16]) were expressed as GST fusion proteins in BL21 (DE3) cells. GST-tagged proteins were isolated by glutathione sepharose (GE Healthcare) using a buffer containing 50 mM Tris–HCl, pH 7.5, 200 mM NaCl, 4% glycerol, 5 mM DTT. Proteins were further purified by either elution with 5–8 mM glutathione (Fisher BioReagents), or on-column cleavage by TEV, followed by anion exchange and size-exclusion chromatography. All proteins were concentrated by ultrafiltration to 3–10 mg ml^{-1}.

Crystallization, data collection and structure determination. The crystals of ASK1–D3 form 1 complex were grown at 4 °C by the hanging-drop vapour diffusion method with 1.0 μl protein-complex sample mixed with an equal volume of reservoir solution containing 6.5% CP-42, 175 mM sodium citrate tribasic dihydrate, 87 mM HEPES sodium pH 7.5 and 26% MPD. Crystals of maximal sizes were obtained and collected after 2 weeks. The heavy-atom derivative ASK1–D3 form 1 crystals were prepared by soaking the native crystals in the presence of 10 mM $K_2Pt(NO_2)_4$ for 4 h. The crystals of the ASK1–D3 form 2 complex were grown at 25 °C by the hanging-drop vapour diffusion method with 1.0 μl protein-complex sample mixed with an equal volume of reservoir solution containing 80 mM Tris–HCl, pH 7.0, 24% MPD, 24% PEG1000, 24% P3350; 15 mM sodium citrate tribasic dihydrate pH 5.6 and 0.5 M 1,6-hexanediol. The crystals of ASK1–D3 form 3 complex were grown at 25 °C by the hanging-drop vapour diffusion method with 2.0 μl protein-complex sample mixed with an equal volume of reservoir solution containing 150 mM Tris–HCl, pH 7.4, 22% MPD, 22% PEG1000, 22% P3350, 15 mM sodium citrate tribasic dihydrate pH 5.6, 0.45 M 1,6-hexanediol, and 5 mM DTT. The crystals of D14–D3-CTH were grown at 4 °C by the hanging-drop vapour diffusion method with 1.0 μl protein-complex sample mixed with an equal volume of reservoir solution containing 0.02 M amino acid mixture (Glu, Ala, Gly, Lys and Ser); imidazole; 0.1 M MES monohydrate, pH 6.5, 40% glycerol and 20% PEG4000. The single anomalous dispersion dataset was collected near the platinum absorption edge ($\lambda = 1.072$ Å). X-ray diffraction data were integrated and scaled with HKL2000 package[37]. Single anomalous dispersion was used to determine the initial phase using PHENIX[38] with a 2.5 Å platinum derivative dataset for ASK1–D3 form 1. Initial structural models were built, refined and rebuilt using COOT[39]

and PHENIX. The final model was built and refined with a native dataset. The crystals of ASK1–D3 form 2 and ASK1–D3 form 3 complexes were determined by molecular replacement using ASK1–D3 form 1 structure as the search model. The D14–D3-CTH structure was determined by molecular replacement using rice D14 structure (PDB 4IH9)[23] as the search model. All structural models were manually built, refined, and rebuilt with PHENIX and COOT.

AlphaScreen luminescence proximity assay. AlphaScreen assays for determining and measuring protein–protein interactions were performed using EnSpire reader (PerkinElmer). GST-tagged D53 or D14 was attached to glutathione AlphaScreen donor beads. His-tagged D14 or D3 was attached to anti-$6 \times$ His conjugated AlphaScreen acceptor beads. The donor and acceptor beads were brought into proximity by the interactions between D14, D53 and ASK1–D3 complex, which were measured with and without GR24 and/or non-tagged proteins at indicated concentrations. When excited by a laser beam of 680 nm, the donor beads emit singlet oxygen that activates thioxene derivatives in the acceptor beads, which then release photons of 520–620 nm as the binding signal. The experiments were conducted with 100–500 nM of D14 or D53 and 1 μM ASK1–D3 complex in the presence of 5 μg/ml donor and acceptor beads in a buffer of 50 mM MES, pH 6.5, 150 mM NaCl, 1 mM DTT and 0.1 mg/ml bovine serum albumin. The results were based on an average of three experiments with standard errors typically <10% of the measurements. Half maximal inhibitory concentration values were determined using nonlinear curve-fitting of graphs generated with Prism 6 (GraphPad).

YLG hydrolysis assay. YLG (TCI America) hydrolysis assays were performed using 1–2 μg of recombinant proteins in a reaction buffer (50 mM MES pH 6.5, 150 mM NaCl and 1 mM DTT) at a 100-μl volume on a 96-well black plate (Greiner). The fluorescence intensity was measured by EnSpire 2300 multilabel plate reader (PerkinElmer) at excitation by 480 nm and detection by 520 nm. Ninety-six-well black half-area plates were covered with Viewdrop III UV plate seals to prevent evaporation. Time-course experiments were performed in 10-s intervals over 50–60 min. Fluorescence data were converted directly to fluorescein concentration using a standard curve. Data generated in Excel were transferred to Prism 6 for graphical analysis and curve-fitting. In all cases in which synthesized peptides (Genscript and Biomatik) were analysed, dimethylsulfoxide (DMSO) was added in equivalent concentration into the reaction.

Size-exclusion chromatography. Purified proteins (20–50 μM) were incubated with 100–200 μM GR24 (Chiralix), or equal amount of acetone as the solvent control, at 4 °C for one hour in 20 mM HEPES, pH 7.0, 150 mM NaCl and 2 mM DTT. The reaction was injected onto a Superdex-200 Increase 10/300 column (GE Healthcare) for analysis at a flow rate of 0.5 ml min^{-1}. The elution fractions (0.5 ml per fraction) were resolved by SDS–PAGE and analysed by Coomassie blue G-250 stain.

Limited proteolytic digestion. One milligram per millilitre of purified ASK1–D3 (or ASK1–D3(ΔCTH)) was incubated at 4 °C for 12 h with increasing amount of trypsin solution containing 0.05 mg ml^{-1} trypsin (Agilent) at the volume ratio of 1:3,000, 1:1,500 and 1:750 in 40 mM Tris–HCl, pH 7.5 and 1 mM DTT. The proteolysis reactions were stopped by fivefold-concentrated SDS–PAGE sample buffer immediately followed by 5 min boiling at 95 °C. Proteins were resolved by SDS–PAGE and Coomassie blue G-250 stain. The resolved bends corresponding to ASK1 and digested D3-CTD were excised and further analysed by N-terminal sequencing (Analytical Core Facility, Tufts Medical School).

Affinity pull-down assay. Pull-down assay was performed using ~20–40 μg of purified GST-tagged proteins as the bait and ~10–25 μg of either His-tagged or non-tagged proteins. Reaction mixtures were incubated with GST beads (GE Healthcare) at 4 °C for 30 min in the reaction buffer with 40 mM Tris–HCl, pH 7, 100 mM NaCl 2 mM DTT and 0.01% bovine serum albumin. After an extensive wash with a buffer containing 40 mM Tris–HCl, pH 7, 250 mM NaCl, 2 mM DTT and 0.01% (v/v) Tween 20, the protein complexes on the beads were either eluted by fivefold-concentrated SDS–PAGE sample buffer or by 5 mM glutathione. All samples were boiled at 95 °C for 5 min and resolved by SDS–PAGE. Proteins were analysed as indicated by Ponceau stain or western blot analysis using specific polyclonal anti-GST antibody and monoclonal anti-His antibody (Sigma). In pull-down assays in the presence or absence of 20–100 μM GR24, acetone was used as solvent control. Inputs samples represent 5–10% of the total reaction.

D53 stability in a reconstituted cell-free system. *Arabidopsis* ecotype Columbia-0 (Col-0) wild-type and *max2-1* mutant seeds were surface-sterilized with 50% (v/v) bleach and 0.1% Triton X-100. After cold treatment at 4 °C for 48 h, seeds were germinated and grown on 0.5 × Murashige and Skoog (MS) medium containing 0.8% agar and 1% sucrose under continuous light at 22 °C. One hundred milligrams of 7-day-old seedling were collected and frozen in liquid nitrogen. Total proteins were extracted after grinding with native protein extraction kit (Minute, Invent Biotechnologies) supplemented with protease inhibitor cocktail (Roche), followed by two sequential centrifugations at 12,000g for 10 min. To monitor protein degradation in the cell-free system, 0.5 μg of purified GST-tagged proteins (either full-length D53, D53-D2 or d53-D2, as indicated) was incubated at 22 °C

in a reaction mixture that contained, at a final volume of 12.5 μl, 1–2 μl of plant extract supplemented with 10 μM GR24, 25 mM Tris–HCl, pH 7.4, 0.625 mM ATP, 5 mM MgCl₂ and 0.5 mM DTT. Where indicated, the proteasome inhibitor MG132 (Calbiochem) was added at a concentration of 100 μM. Reactions were terminated at the indicated times by the addition of fivefold-concentrated sample buffer. Boiled samples were resolved via SDS–PAGE, and proteins were visualized using western blot and polyclonal anti-GST antibodies (Sigma).

Plant growth conditions. *Nicotiana benthamiana* plants were grown on F2 compost (Levington Horticulture) pre-treated with 0.2 g L-1 Intercept (Everris). Glasshouse conditions as follows: 16 h light:8 h darkness, minimum irradiance 88 W per m², shading implemented at 500 W per m² and cooling implemented at 31 °C. Humidity and temperature were determined by the ambient conditions.

Cloning and plant transformation. All constructs were cloned using Multi-site Gateway (Invitrogen). A previously generated CaMV 35S promoter and SMXL7–YFP expression vector were used[25]. D14, D14-CTH and CTH-NSL sequences were synthesized and cloned into pDONR221. The mCherry and YFP fluorescent tags were cloned into the pDONR P2-P3R and all ENTRY vectors were then recombined in the relevant combinations in pH7m24GW (https://gateway.psb.ugent.be/). *Agrobacterium tumefaciens* strain GV3101 was transformed using standard electroporation procedure.

Assays for transient gene expression mediated by *A. tumefaciens* in *N. benthamiana*. *A. tumefaciens* (strain GV3101) carrying the desired transfer DNA construct was grown overnight at 28 °C with the appropriate antibiotics. Cells were collected by centrifugation at 8,000*g* and resuspended in agro-infiltration medium with 5 mM MES, 10 mM MgCl2, pH 5.6, before syringe infiltration into leaves of 3–4-week-old *N. benthamiana* plants. Bacteria carrying each construct were infiltrated at a final OD₆₀₀ nm of 0.4. Leaves were detached 48 h post-infiltration for confocal imaging.

Confocal microscopy. All confocal images were captured on a Leica SP8 laser scanning confocal using a W Plan-Apochromat 20× 1.0 numerical aperture objective (Zeiss). Detection wavelengths: 520–540 for SMXL–YFP, and 600–620 for mCherry-tagged proteins. The pinhole was set to one airy unit for all nuclei. Detection gain and laser power were kept constant between t_1 and t_0 for intensity

comparison and the same settings were used for all nuclei of the same construct combination. Laser power was adjusted across construct combinations as necessary to account for differences in expression level and avoid signal saturation.

SMXL7 quantification. Two days after infiltration, leaves were infiltrated with *A. thaliana* salt (ATS) with 0.1% v/v acetone for the mock or 10 μM GR24, 0.1% v/v acetone for the treated samples. Between 7 and 13 nuclei expressing 35S:SMLX7–YFP alone or in combination with either 35S:D14–mCherry, 35:D14-CTH–mCherry or 35:CTH-NLS–mCherry were located and imaged at time 0. The same nuclei were then imaged again using identical settings after 120 min for each construct combination and treatment. Areas of interest were drawn around each nucleus in ImageJ version 2.0.0 and the mean signal intensity was recorded. The ratio of the means at 0 min and 120 min post-treatment was used to compute relative fluorescence at t_1, which is expressed as the percentage of change in the level of SMXL7. Distributions for each combination and treatment condition were compared using a two-tailed Student's *t*-test with Bonferroni correction.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Structural coordinates and structural factors have been deposited in the RCSB Protein Data Bank under accession numbers 6BRO, 6BRP, 6BRQ and 6BRT. Uncropped gels and blots are available in the Supplementary Information. All other data are available from the corresponding author upon reasonable request.

36. Su, Y., Zou, Z., Feng, S., Zhou, P. & Cao, L. The acidity of protein fusion partners predominantly determines the efficacy to improve the solubility of the target proteins expressed in *Escherichia coli*. *J. Biotechnol.* **129**, 373–382 (2007).
37. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
38. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

a

LRR20 α-helix
(D3-CTH)

Rice	M F T E M R A E S W L R F E V Q L N S R Q I D D	720
Arabidopsis	M S T E M R V G S C S R F E D Q L N S R N I I D	693
Castor	M S T E M R V G S C S R F E D A L N R R H I V D	695
Poplar	T C T E M R V G S C S R F E D A L N R R Q I L D	694
Grapevine	M S T E M R I D S C S R F E D A L N R R R I L D	712
Cucumber	M S T E M R A G S C S R F E A A L N S R Q I P D	715
MonkeyFlower	M S T E M R S D S C S R F E A A L N R R Q I S D	713
Tobacco	M S T E M R A D S L S R F E A A L N R R P I S D	724
Medicago	M S T E M R V G S C I R F E D A L N R R Q I C D	711
Pea	M S T E M R V G S C S R F E D A L N R R I I C D	708
Soybean	M S T E M R V G S C S R F E D A L N R R R I C D	711
Maize	M N T E M R A E S W L R F E V Q L N N R L I E D	705
Sorghum	M N T E M R A E S W L R F E N Q L N I R L I E D	700
Moss	T T T E L R S V S C Q R F E A L V A K R G F P D	698

b

c

d

Extended Data Fig. 1 | Conservation and conformation of D3 C-terminal α-helix. a, Sequence alignment of the C-terminal regions of 14 orthologues of MAX2 or D3. Highly conserved residues are coloured in blue. **b**, Electrostatic-potential surface map of D3 with CTH shown in cartoon representation (orange). The C terminus aspartic acid residue (Asp720) is anchored to a positively charged pocket. **c**, Close-up view of the D3 extreme C-terminal residue (Asp720) and its interacting residues in D3 and ASK1. **d**, Electron densities of the D3-CTH region in two different crystal forms, adopting either a regular helical conformation (left) or an extended conformation (right).

```
Rice         MAEEEEVEEGRSSSSAILDLPEPLLHILSFLTDVRSRHRAALACGRMRAAERATRSELSLRGDPRSPGFLFLSH-AFRFPALEHLDLSLVSPWGHPLLSSVPPCGGGGGGA 111
Arabidopsis  -----------MASTTLSDLPDVILSTISSLVSDSRARNSLSLVSHKFLALERSTRSHLTIRGNA--RDLSLV--PD-CFRSISHLDLSFLSPWGHTLLA----------- 84
Castor       -------MTITTTATINDLPDVILSNIIASISDTRTRNSLSLVNRKFLTLERTRTSLTLRGNA--RDLYMI--PT-CFRSVTHLDLSLLSPWGHSLLA----------- 87
Poplar       --------MAATMNDLPDVILSIIFSSVSDTRTRNSLSLVNRKFLLERSTRTSLTLRGKA--RDIYMI--PT-CFRSVTHLDLSLLSPWGRSDLL----------- 83
Grapevine    -----MAGAAAGGATTIHDLPDAILSSIASVTDTRARNAAALVCRKWLVLERGTRTSLTLRGNVVHNNLYMI--PT-CFRAVTHLDLSLLSPWGHSLI----------- 91
Cucumber     -----MAARTSINDLPDVLVSNILALVSDTRTRNSLSLVCRKFLSLERTRTFSLSLRGNA--RDLYGI--PT-CFRSVTHLDLSLLSPWGHAFLC----------- 84
MonkeyFlower -----MAAPPTSGTTLNDLPDVILSNIIAAVVDVRGRNSAALVCSKWYALERATRSSITLRANL-RDLFMV--PN-CFRSVSHLDSLLSPWGHPLTS----------- 89
Tobacco      -----MATATQLTCSTIINDLPDVILSNIIAAISDVRSCNSAALVSRKWLVLERSTRVSLTLRGNV--RDLFML--PT-CFRSITHLDLSLISPWGHPLLSPVSG----- 95
Medicago     -------MVGNNSATTVSHLPEEILSKVFTGITDTRTRNSLSLVCHSFFKLERKTRLSLTLRGNA--RDLYRI--PT-SFTNVTHLDVSLLSPWGHALFC--SP------ 90
Pea          ------MVD----TTVAHLPEEILSQVFAAITDTRTRNSLSLVCRSFFRLERKTRVSLTLRGNA--RDLYRI--PT-SFAHVTNLDVSLLSPWGHALFC--SP------ 86
Soybean      ---------MGDGSIVGHLPEEILLNVFSAVSDTRTRNALSLVSWSFYHLERKTRLSLTLRGNA--RDLHLI--PT-SFKHVTHLDSLSFLSPWGHALFC--SS------ 87
Maize        MAEEDAAA----AAGSPILDLPEPLLHILGFLDDARSRHRAALACHRLLAAERATRAALSLRGDPRSNAFLFL-RPTFCFPALERLDLSLVSPWGHPFLSSAAP--SADAVA 106
Sorghum      MAED---A---AAGSPLLDLPEPLLHILGFLTDARSRHRVALACHRLLAAERATRAALSLRGDPRSDVFLYLIRPTFCFPALERLDLSLVSPWGHPFLSSAAP--SADAVA 104
Moss         ---MQSSPEAAPAGTHISDLPSVILTNIIAYVSNPRVRNCISLACRDWYFIERQTRTELSLRGNI----CVMHELPT-CFQQICTLDLSQCSPWGSSLF---------- 91
```

```
Rice         PSASSSSGMNVYHPEAISEQNAFIAARLAGCFPAVTSLAVYCRDPTTLANLT-PHW-QASLRRVKLVRWHQR----PPTLPDGADLEPLLETCA-ALRELDLSEFYCWTEDV 216
Arabidopsis  ----------SLPIDHQNLLALRLKFCFPFVESLNVYTRSPSSLELL-PQW--PRIRHIKLLRWHQR----ASQIPTGGDFVPIFEHCGGFLESLDLSNFYHWTEDL 175
Castor       -----------SSLPSDPLLLAHRLGIAFPLVTSLTVYARSPCTLHVLI-PQW--PLLSHVKLIRWHQR----PSSSQLGADFVPLFEQC-KLLSCLDLSSFYYWTEDV 177
Poplar       -----------STASSVPFLLAQRLRLAFPLVTSLTVYARSPSTLHILL-PQW--PNLSHVKLIRWHPR----SSSPHLGNDVVPLFEHC-QALSSIDLSFYYWTEDI 173
Grapevine    -----------SPSSDPMLLAHLLRHAFPMVTSLTVYARTPATLQLLA-PQW--PNLTHIKLVKWHQR----SPSA-LGSDFDPILRHC-TSLTSVDLSNFYYWTEDL 179
Cucumber     -----------SSP--DPDLLAHRLRGLFPLVTSLTVYARTPTTLQILA-RQW--PELRHVKLVRWHQR----PQSAP-GEDLAPIFEHC-RSLSTLDLSEFYYWIEDI 171
MonkeyFlower -----------ASDPALIAHLLRLAFPSLTSLTFYARNPSTIQLVA-SQW--HNVEHLKLVRWHQR----PQVETAGDELKILFSECGKNLNSLDLSAFYCWTDDV 177
Tobacco      -----------AAAGADPSLIAHLLRHAFPSVTSLVVYTRHPFTLRLLP-PLW--PHLKETKLVRWHQR----PQLAT-GDEFNMLFENCPQ-LKSVDLSTFYCWTDDI 184
Medicago     -----------AGN--DSPLLAQRLRNTFPRVTSLTVYYRDPHTLHLLLFNHW--PELRDVRLVRWHQR----POGLQPGSDFDALFSRC-RSITSLDLSSFYHWPEDL 179
Pea          -----------ATA--DSPSLAQRLRNAFPRVTSLTVYVRDPTLHLLHSHW--PELRDVRLVRWHQR----PPDLQPGSDFAALFSRC-RSITSLDLSSFYHWPEDL 175
Soybean      -----------SATVGHQSLLAQHLRAAFPRVTSLAIYARDPYTLRLLLLSAW--PELSAVKLVRWHQR----PPTSANEADFAELFKKC-RSLASLDLSSFYHWTEDI 178
Maize        PSVTA--------EEVAEQNAFIAARLAYCFPAVSSLAVYCRDPTTLASLT-PHW-RSGLRSVKLVRWHQR----PPGLDAGADLEPLLEDCP-ALRTLDLSEFYCWTEDI 202
Sorghum      PAVAA--------EEVAGQNALIAARLAYCFPAVSLAVYCRDPTTLASLT-PHW-RSRLRSVKLVRWHQR----PPGLDAGADLEPLLGDCP-ALRALDLSEFYCWTEDI 200
Moss         ---------------QSTQNGEEIGNCLRIGFPNVVNLTVYVRDALDIQMVA---WIWPDLEIVKLVRWHPRAMESSEADDLGNEIEGLLSACKR-LKSLDLSKFYCWTEDI 184
```

```
Rice         VRALTTHPSATAALTHLDLGLAAATDGFKSSELGPIAASCPNLRKLVAPCLFNPRFSDCVGDDALLSLATSCPRLTVLRLSEPFEAAANI--------QREEAAITVAGLVA 320
Arabidopsis  PPVLLRYADVAARLTRLDLLTASFTEGYKSSEIVSITKSCPNLKTFRVACTFDPRYFEFVGDETLSAVATSSPKLTLLHMVDTASLANPRAIP---GTEAGDSAVTAGTLIE 284
Castor       PPVLEAYSDVSKSLTCLDLLTVSLTDGFKSDEIKVITAACTNLTKFLVACMFDPSYLGFTGDETLLAVAANCPKLSVLHLVDTSSLGNIRSDPEDEGYSGDDARVSVNGLVD 289
Poplar       PPVLQAYPSVSKALTCLDLLTVSLTDGFKSEEIQAITAACPSLTRFLLVCIFDPSYFGCVGDETLLAIVANCPRLRVLHLVDRASLGSTRGEPEDDGYTREDARITKVGLVD 285
Grapevine    PPALQAHPATAAALTRLDLMTLSFAEGFKSHEILAITAACPNLQQLIACTFDPRYIGFVGDEAIVAIASNCPGLTVLHLADTASLSNGRGDPEEEGFSSEDAGISTTALSG 291
Cucumber     PPVLVANPLTARSISKLNLMTTSLTDGFKSTDIETITEACPNLSQLLMACTFDPSYLGFVGDETLSAIATNCPRLSLLHLADTSTLASVRGDPSADGFTPEDARISTATLIE 283
MonkeyFlower PPALESCPSVSSNLTRLNLLNPSFSEGFKSDEIKSITKACPNLKEFRAACMFDPRYMGSVGDEALVSVSVNCPKLSLLHLADTSALSNTRGDPEHDGFTQEDAKINVATLIE 289
Tobacco      PTALESHPMVASNLTSFNLLNSSFPEGFKSDEIKVITQACPNLKEFKVACMFDPRYIGFVGDEGLVCIATNCSKLVLHLADTSALSNSRGGPNDEGFTEEDAKISVGTLIE 296
Medicago     PPVLAENTTTAASLRRLNLLTTSFTEGFKSNQIESITSSCPNLEHFLVACTFDPRYIGFVGDETLLAVASNCPKLKLLHMADTSSFSNRR---EEEG--VEDARVSRATLVA 286
Pea          PPVLAANAAAAISLRRLNLLTT-FTEGFKSNQIESITSSCPNLEHLLVACTFDPRCIGFVGDETLLAIASNCPKLSLLHMADTSSFSNRR---EEEG--GEDASVSRATLLA 281
Soybean      PKVLAANPISAATLRRLNLLTTSLPEGFKAHEIESITSSCPNLEHFLVAPCVFNPRYIGFVSDDTLVAIPSNCPKLSLLHLADTSSFLNRR---EDEGFDGEDASVSRAALLT 287
Maize        EPALAAHPAAAAILTELDLGLAGATDGFHAAELEAIAGSCPSLQKLVAPCVFNPRYIDFVSDDALLTIAARCPKLAILRLREPFEPAATG-------QREDAAITVAGLVS 306
Sorghum      QPALAAHPAAAALTELDLGLAGATDGFYAAELGAIARFCPNLRKLVAPCTFNPRYIDFVSDDALLTIATSCPKLSIIRLREPFELAATS-------QREDAAITVAGLVS 304
Moss         PPALRAGASTAANLRVLNLLKLS-PNGFKAQEVGAITSSCFNLREFYILCDFDHRLLDSVGDEALLSIATNCPLLKVLHLVDYNEWSAVSDDPNQDAFAAEDSSLSRQGLEA 295
```

```
Rice         FFAALPALEDFTMDLQHNVLEAAPAMEALARRCPRIKFLTLGSFQGLCKA-SWLHLDGVAVCGGLESLYMKNCQDLTDASLAAIGRGCRRLAKFGIHGCDLVTSAGIRRLAF 431
Arabidopsis  VFSGLPNLEELVLDVGKDVKHSGVALEALNSKCKKLRVLKLGQFQGVCSATEWRLDGVALCGGLQSLSIKNSGDLTDMGLVAIGRGCCKLTTFEIQGCENVTVDGLRTMVS 396
Castor       FFSGLPLLEELVLDVRVCKNVRDSFVALEALNSRCPKLKLELVQFHGVCMAVE-SQLDGVALCSGLKSLSIKKCADLTDMGLIEIARGCCRLAKFEVEGCKITMKGLRTMAS 400
Poplar       FFTGLPLLQELVLDFYQNVRDSALALEALHSKCPELKLLKLGQFHGICMAID-SQLDGVALCSGLVSLTIKNSADLTDMGLIEIAARGCSKLAKFEVEGCKITMKGMRTMAS 397
Grapevine    LFSGLPLLEDLVLDVAKNVRDSGATELMLSNSRCPKLRVLKLGHFHGLCLAIG-SQLDGVALCQGLESLSIKNSADLTDMGLIAIARGCSKLAKFEIHGCKKVTWKGISTMAC 402
Cucumber     LFSGLPLLEDLVLDVAKNVRDSGPALEVLNTKCRRLSLKLGQFHGLCMAID-SRLDGIALCGGLESLSITNCADLTNMRLIEVGRGCVRLSKFEVKGCKITVKGLRTMVS 394
MonkeyFlower VFSGLPLLEELVLDVCNNVRESGPALEVLNSKCPKLRSLKLCQFHGVSLPIE-SKLDGIALCHGLESLSIRNVGDLTDMGLIAIGRGCCRLSVFEVHGCKNITVRGMRTLAS 400
Tobacco      VFSGLPLLEELVLDVCNNVRDTGPALEILNKKCPQLRSLKLGQIHGISMPIE-SKLDGVALCQGLQSLSIRNVGDLNDMGLIAIGRGCCRLAKFEIQGCKKITMRGMRTLAS 407
Medicago     FTGLPLLEELVLDVCKNVTETSFALEMLSSKCPNLKVVKLGQFQGICLAIG-SRLDGIALCHGLQSLSVNTCGDLDDMGLIEIGRGCSRLVRFEIQGCKLVTEKGLRTMAC 397
Pea          LFSGLPLLEELVLDVCKNVSESSFAFEMLSSKCPNLKVVKLGHFQGICLAIG-SRLDGIALCHGLQSLSVICCGDLDDMGLIEIGRGCSRLVRFEIQGCKLVTEKGLKTMTC 392
Soybean      LFSGLPLLEELVLDVCKNVRESSFALEVLGSKCPNLRVLKLGQFQGICLAFG-SRLDGIALCHGLSVGNCADLDDMGLIEIARGCSRLVRFELQGCRLVTERGLRTMAC 398
Maize        FFAALPALEDFTLDLRHNVLETAPAMEALARRCPRIKFLTLGGFQGLCKA-SWLHLDGVAVCGSLESLCIKGCLDLTDASLAAIGRGCGRLAKFAIHGCDLVTPAGIRRLAT 417
Sorghum      FFAALPALEDFTLDLRHNVLETAPAMEALARRCPRIKFLTLGGFQGLCKA-SWLHLDGVAVCGSLESLCIKGCLDLTDASLAAIGRGCGRLAKFAIHGCDLVTSVGIRRLAT 415
Moss         MFKALPHLEDLVFYLSQNLRDSGAPFEILASSCKKLRSLKLSNFLGVC--GGPHPDGIALCHALQELRLKNCGDLTDDALKAISVGCPKLSKLGLRQCKSITKEGLHACVK 404
```

```
Rice         TLRPTLKEVTVLHCRLLHTAECLTALSPIRDRIESLEINCVWNTTEQPCSVANGTTT-ECDPEDDELG-EV--------------YESAAKKCRYMEFDDL-- 516
Arabidopsis  LRSKTLTDVRISCCKNLDTASLKAIEPICDRIKRLHIDCVWSGSED-EEVEGRVETSEADHEEEDDG-------------YERSQKRCKYSFEEHCS 481
Castor       LLHKTLVEVKISACKNLDAVASLRALEPIRQRIERLHIDCMWNSLQE-EDNYGGN--HSFDLNEILFGSD--------EH-E--------YSSRNKRIKYSKD---- 483
Poplar       LLHKTLIEVKISCCKNLNAVASLRSLEPIQGRIERLHFDCVWEGLEE-D---GGI--LCFDLNEGLCQSV--------EH-E--------YGSKRKKSKYSSDPD-SS 480
Grapevine    LRRSTLVEFKISCCKNLDAVSALRGLEPIRDRIQRLHIDCIWDKSEQFEDSEEAILAHSFDLNELQRSIPSQDDDRFWDH--E--------ASIKKKKRKYTTDLD---- 499
Cucumber     LLKRTLVDVKISCCENLNTKASLRALEPIQDRISRLHVDCVWKDVEECELEYDTASSSNIDPDEVDELTLPSHNADTSSST--DGLLEDGNYGGYTRKRKRSYSTDAD-CS 503
MonkeyFlower LLRRTLIDVRISCCKNLGAEQSLKALEPIQDRIESLHIDCIWD--------DGT--ESGFDLNDSFDQGEVSNDENHFESRGYNDYY---DAATIENPSKKCKYSYDLN-SS 498
Tobacco      LLKKTLVDVKISCCKNLGASSSLKALEPIQERIQKLHIDCVWDSVKEFENIDGY--GYGLDLNRS-DGGEASSN---FAGSG--DTFGCEEEAYMLQQKKRCKFTCDLN-CL 510
Medicago     LLRRTLIDVKVASCVNLDAAATLRALEPIRDRIERLHLDCIW---KESDNFGQGLF--NFDLNTLDEL-NGSELMDCFGGEECGE------DTSMRKRQRCEYGLEAD-DL 495
Pea          LLRRTLIDVKVASCVNLDAAATLRALEPIRDRIERLHLDCVW--KESDNLGHSFL--NFDLNASAEL-NESELMECFGGEEYGE------DTSRRKRQRCEYGFEDD-DS 490
Soybean      LLARTLIDVKVSCCVNLDTAATLRALEPIRERIERLHVDCVWNGLKESDGLGHGFL--NFDLNGLDEPGDGGELMDYFGGGEC-E------NATKRKRQRCEYEMGVH-DS 499
Maize        ALRPTIKEVSILNCRLLDTAACLTALSPIRDRIESLEISCVWKEVEQPESVANGIAG---CNHEDDDLGGEVS------------YESASKKCRYMELDDL-- 503
Sorghum      ALRPTIKEVSILHCRLLDTATCLTALSPIRDRIESLEVSCVWKEVEQPESVANGTTG--CDHEDDDLG-EVT--------------YESASKKCRYMELDDL-- 500
Moss         NLSHTLKDVEIAGCKLLPTAMTLKALEPIQVTVKNLHLDCVWDEGILAQEASAARTQSTVDSLNHEQSAR-----------SMGPGGTQQLSVPGAKKIQTSS----- 496
```

```
Rice         ----------------GSWEMLRSLSLWFSAGQLLSPLISAGLDSCPVLEEISIKVEGDCRTCPRPAPRTIFGLSDLAGFPVLAKMKLDLSEAVGYALTAPTGQMDLSLWE 611
Arabidopsis  TSDVNGFCSED------RVWEKLEYLSLWINVGEFLTPLPMTGLDDCPNLEEIRIKIEGDCRGKRRPAEPE-FGLSCLALYPKLSKMQLDCGDTIGFALTAPPMQMDLSLWE 586
Castor       GFCMQN----NG--VWSNSWDNLKCLSLWIGVGELLTPLPMAGLEDCPSLEEIQIRVEGDCRGRHKLSQRA-FGLSCLAHYPRLSKMQLDCSDTIGFALTAPSGQMDLSLWE 588
Poplar       SSCMQS----NGNGMFSKSWDRLKYLSLWIGAGVLLTPLPMAGLYDCPNLEEIRIKVEGDCRTGHKPSQRE-FGLSCLAHYPRLSKMQLDCSDTIGFALTAPSGQMDLSLWE 587
Grapevine    ASYEQN----NGNGICSKTWERLRCLSLWIGVGELLPPLAKAGLDDCPVLEEIQIKVEGDCRERSKPSQ-P-FGLSSLMRYPRLSKMKLDCGDTIGYALTAPSGQTDLSTWE 605
Cucumber     LS-IQC----NGNDLWGKRWDRLEYLSLWIGVGDFLSPLETVGLDDCPVLQEIQIKVEGDCPRRHKPMD-T-FGLSIGQYPQLAKMKLDCSDTTGYALTCPSGQMDLTLWE 608
MonkeyFlower -SDLNGY--ENGNVYGERTWERLKYLSLWIAVGQLLNPLTNAGLENCPNLEEIRIKIEGDCRELPKPSERE-FGLSNLVNYPRLSKMKLDCGETIGYAHTAPSGQMDLSLWE 606
Tobacco      YEEVNGH--GNG--YSGRSWDRLQCLSLWIGVGELLTPLTVAGLEDCPNLEEIRIVEGDCRLWSKPSERA-FGLSTLLLYPKLSKMHLDCGDTIGYAHTAPSGQMDLSLWE 617
Medicago     FVQSNGN--GNGNGYYGYSWDRLEYLSLWIKVGELLTQLPVAGLEDCPNLEEIRIKVEGDCRGQPKPAVRE-FGLSILACYPQLSKMQLDCGDTKGYVYTAPSGQLDLSWWE 604
Pea          FVHSNGNSSGNDNGYSCNSWESLHYLSLWIKVGDLLTQLPAAGLEDCPNLEEIRIKMEGDCRGQPKPAVSE-FGLSILTCYPQLSKMQLDCGDTRGYVYTAPSGQMDLSLWE 601
Soybean      FLQSNGN-----GFCGKSWDKLSLHYLSLWIKVGDLLTPLPVAGLEDCPVLEEIRIKVEGDCRGQPKPAESE-FGLSILACYPQLLKMQLDCGDTIGYAHTAPSGQMDLSLWE 604
Maize        -----------VSWEMLRSLSLWFPAGEVLSPLISAGLDSCPVLEEISIKVEGDCRTCARPGP--LFGLSDLAGFPVLAKMKLDLSEAVGYALTAPAGQMDLSLWE 596
Sorghum      -----------VSWEMLRSLSLWFPAGEVLSPLISAGLDSCPVLEEISIKVEGDCR--ARPGP--FFGLRYLAGFPVLAKMKLDLSEAVGYALTAPAGQMDLSLWE 591
Moss         -----------GKVWKSLESLSLWIPVGEVISPLAAMGLEECPALHELKLKVEGDGRLLRKPST-QGWGINSFGRYPKLEKVELDLSEVTGFSLSAPKGFTDLSSWE 591
```

```
Rice         RFYLHGIESLQTLYELDYWPPQPDKDVHHRSLTLPAVGLIQRCVGLRKLFIHGTTHEHFMTFFLSIPNLRDMQLREDYYPAPENDLMFTEMRAESWLRFEVQLNSRQIDD 720
Arabidopsis  RFFLTGIGSL-SLSELDYWPPQDRDVNQRSLSLPGAGLLQECLTLRKLFIHGTTHEHFMNFLLRIPNLRDVRLREDYYPAPEND-MSTEMRVGSCSRFEDQLNSRNIID 693
Castor       RFFLNGIGSL-SIIELDYWPPQPDRDVNQRSLSLPGAGLLAQCLALRKLFIHGTAHEHFMMFLLRIPNLRDVQLREDYYPAPEND-TCTEMRVGSCSRFEDALNRRQIID 695
Poplar       RFFLNGIGNL-SLIIELDYWPPQPDRDVNQRSLSLPGAGLLAECLAMRKLFIHGTTHEHFIMFLLRIPNLRDVQLREDYYPAPEND-MSTEMRAGSCSRFEEALNRRQIID 694
Grapevine    RFYLNGIKNL-TLNELDYWPPQPDKDVNHRSLSLPSAGLLAECVTLRKLFIHGTAHEHFMTFLLAIPNLRDVRLREDYYPAPEND-MSTEMRIDSCSRFEDALNRRRIILD 712
Cucumber     RFYLNGIGSL-GLTELDYWPPQPDRSFNQRSLSHPAAGLLAECLTLRKLFIHGTAYEHFMNFLLNIPYLRDVQLREDYYPAPEND-MSTEMRAGSCSRFEAALNRRQIPD 715
MonkeyFlower RFYLIGIGHL-SLRELDYWPPQPDKDVNQRSLSLPAAGLLQECFGLRKLFIHGTTHEHFMMFLLRIPDLRDVQLREDYYPAPEND-MSTEMRSDSCSRFEAALNRRQISD 713
Tobacco      RFYLFGIGNL-SLRELDYWPPQPDRDVNQRCLSLPAAGLLQECVTLRKLFILGFAHEHFMNFLLRIPDLRDVQLREDYYPAPEND-MSTEMRADSCSLRRFEAALNRRPISD 724
Medicago     RFLLNGIGSL-SLNELHYWPPQPDEDVNQRSLSLPAAGLLQECYTLRKLFIHGTTHEHFMNYFLKIPNLRDVQLREDYYPAPEND-MSTEMRVGSCIRFEDALNRRQICD 711
Pea          RFFLNGIGSL-SLNELHYWPPQPDEDVNQRSLSLPAAGLLQECYTLRKLFIHGTTHEHFMNFFLKIPNLRDVQLREDYYPAPEND-MSTEMRVGSCSRFEDALNRRICD 708
Soybean      RFLLNGIGSL-SLGELHYWPPQPDEDVNQRSVSLPAAGLLQECYTLRKLFIHGTTHEHFMNFFLKIPNLRDVQLREDYYPAPEND-MSTEMRVGSCSRFEDALNRRRICD 711
Maize        RFYLQGIDSLMTLYELDYWPPQDKEVNQRSLTLPAVGLLQGCVGLRKLFVHGTTHEHFLTFFLKVPNLRDMQLREDYYPAPESDMMNTEMRAESWLRFEVQLNNRLIED 705
Sorghum      RFYLDGIDSLMTLYELDYWPPQDKEVNQRSLTLPAVGLLQGCVGLRKLFVHGTTHEHFLTFFLKVPNLRDMQLREDYYPAPESDMMNTEMRAESWLRFENQLNILIED 700
Moss         RHYLVGINELL-LTELDYWPPSDKEVNRRAISLPGAGLLSLCSKLRKLFVHGTAHEHFLNMITGCRCLRDVQLRGDYYPAPEQE-TTTELRSVSCQRFEALVAKRGFPD 698
```

Extended Data Fig. 2 | Sequence alignment and analysis of selected orthologues of D3 or MAX2. Orthologues of D3 or MAX2 are selected and aligned from rice (*O. sativa*) (accession XP_015643693), *A. thaliana* (NP_565979), castor (*Ricinus communis*) (XP_002528551), poplar (*Populus trichocarpa*) (XP_002320412), grapevine (*Vitis vinifera*) (XP_010657042), cucumber (*Cucumis sativus*) (XP_004137031), monkey flower (*Erythranthe guttata*) (XP_012832933), tobacco (*Nicotiana sylvestris*) (XP_009757168), medicago (*Medicago truncatula*) (XP_003607592), pea (*P. sativum*) (ABD67495), soybean (*Glycine max*) (XP_003540983), maize (*Zea mays*) (XP_020394883), sorghum (*Sorghum bicolor*) (XP_002436499) and moss (*Physcomitrella patens*) (XP_024400746). The non-conserved region designed to be truncated by TEV cleavage during recombinant D3 purification is underlined in green.

a

R461 →
S518 →

b

N474 →
L516 →

CTH →

c

d

	ASK1-D3	ASK1-D3 -ΔCTH	
25kD			— D3-CTD
20kD			— ASK1
			— D3-CTD-ΔCTH
Trypsin	− +	− +	
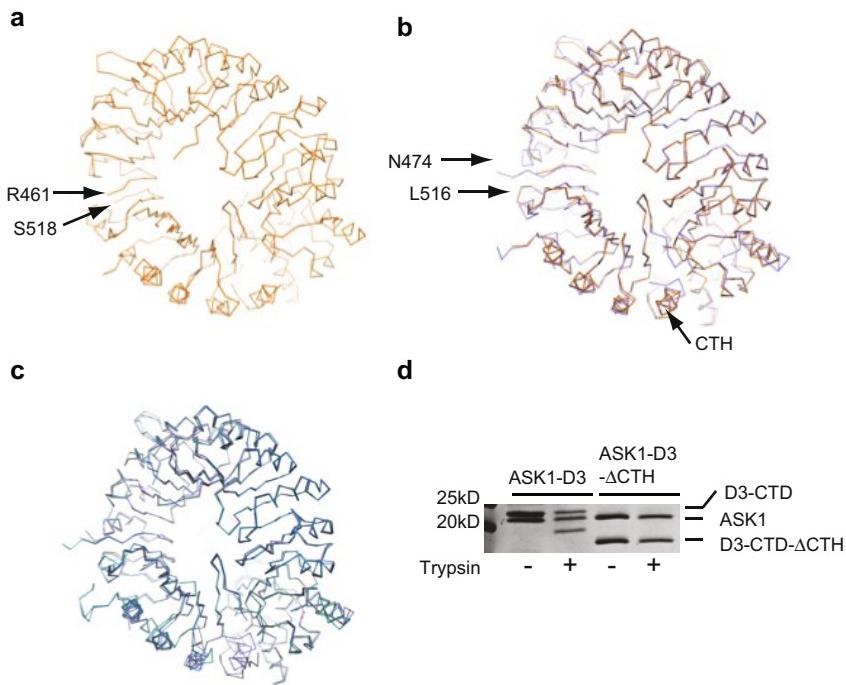
Extended Data Fig. 3 | Comparison of D3–ASK1 structures. a, Top view of ASK1–D3 crystal structure (orange) based on PDB 5HYW. Red arrows indicate a gap in the polypeptide model. Note that PDB 5HYW has a polypeptide register error ranging from amino acid 373 to 473 before the gap. **b,** Superposition of ASK1–D3 determined in this study (light blue) with PDB 5HYW. The region truncated by design ranges from N474 to L516, which are indicated by red arrows. **c,** Superposition of all three crystal forms of ASK1–D3 determined in the current study. **d,** Limited trypsin digestion assay of ASK1–D3 and ASK1–D3(ΔCTH). Proteins were resolved by SDS–PAGE followed by Coomassie blue stain, focusing on D3 C-terminal domain. The experiment was repeated three times.
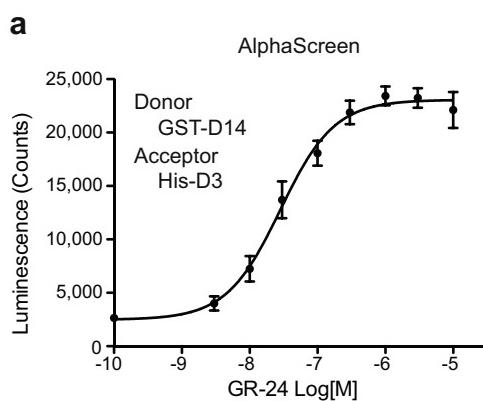
Extended Data Fig. 4 | Established binding between D3 and D14.
a, AlphaScreen assay measuring direct interaction between GST–D14 and His–D3 in response to increasing amounts of GR24 (mean ± s.d. of biological triplicates). **b**, The binding interface between CLIM-bound D14 (magenta) and the LRR domain of D3 (blue) (PDB 5HZG). The last four LRRs are labelled, and D3-CTH in LRR20 is coloured in orange.
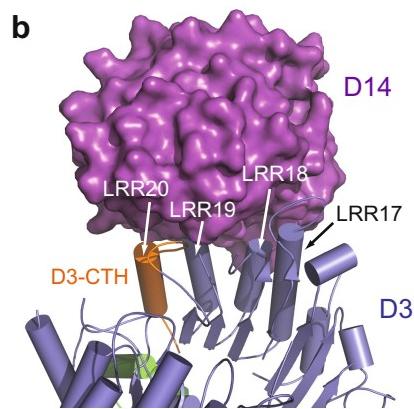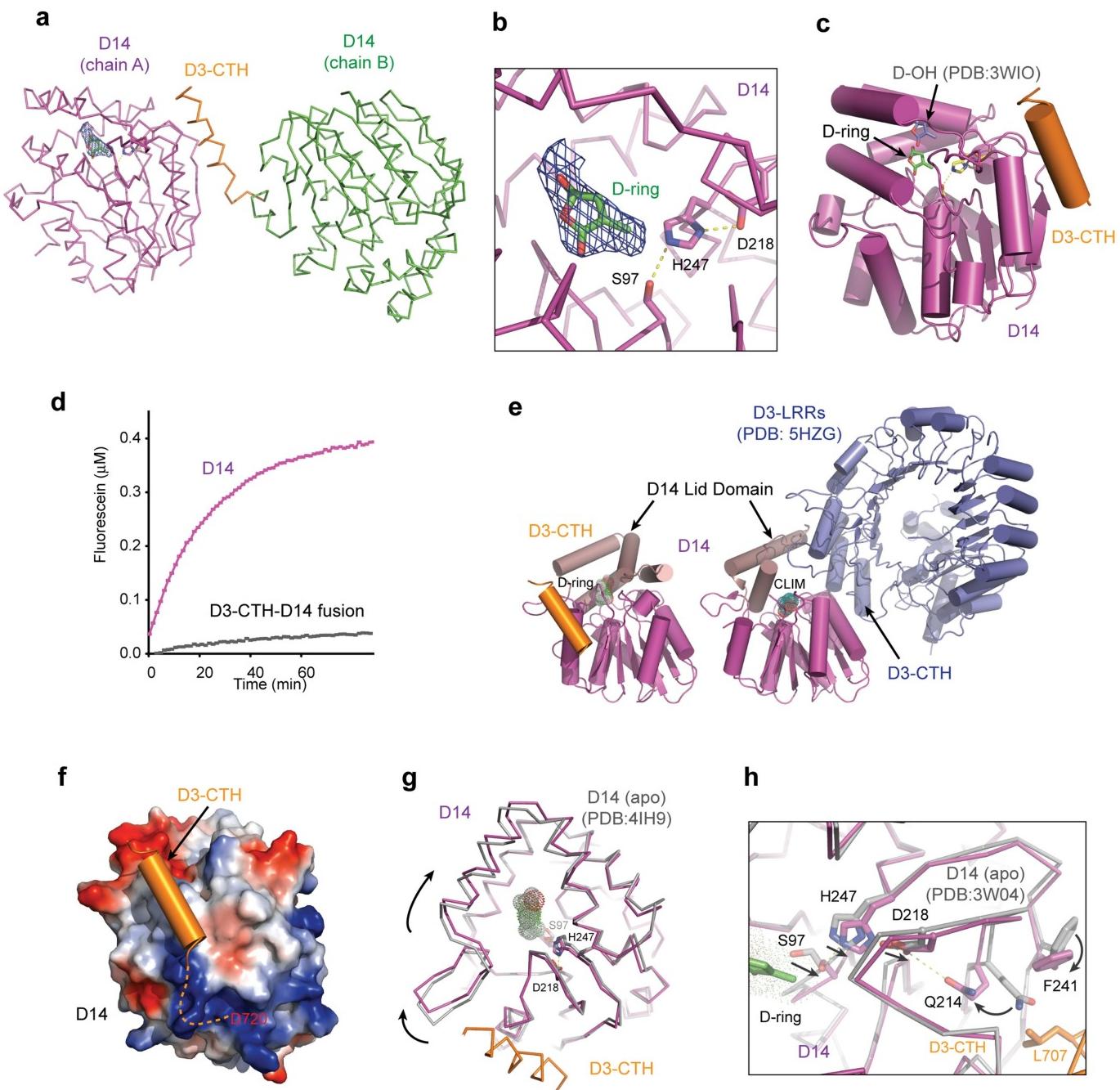
Extended Data Fig. 5 | Structural analysis of D3-CTH–D14–GR24 complex. a, Packing of two D14 molecules that are N-terminally fused with D3-CTH. The D3-CTH region in chain A is omitted. The GR24 D ring (sticks) is shown together with the surround $2F_o - F_c$ electron-density map calculated before the compound modelled in and contoured at 0.8σ. **b**, A close-up view of the GR24 D ring (sticks, green) and its electron density, calculated as in **a**. **c**, Overall structure of D14 (magenta) bound to D3-CTH (orange), with a GR24 D ring (green sticks). The GR24 hydrolysis product D-OH (cyan sticks)—revealed in the D14-D-OH structure (PDB 3WIO)—is shown on the basis of superposition analysis. **d**, Kinetics of YLG hydrolysis by free D14 and D14 fused to D3-CTH. Experiment repeated three times. **e**, Comparison of the interface that D14 (magenta and brown) makes upon binding to D3-CTH (orange) versus upon binding to ASK1–D3 (blue). The lid domain (brown) of D14 adopts open and closed conformation upon binding to D3-CTH and ASK1–D3, respectively. **f**, Electrostatic-potential surface map of D14 bound to D3-CTH (orange). The dashed line indicates the C-terminal region of D3 that would otherwise be free, if D3-CTH were not fused to another copy of D14 in the crystal. **g**, Conformational changes in the lid domain of D14, induced by D3-CTH binding, as revealed by superposition analysis between D3-CTH-bound (magenta) and apo D14 (grey, PDB 4IH9). Arrows indicate the rotation of the lid domain of D14, induced by D3-CTH (orange), relative to the catalytic triad shown in sticks. **h**, Superposition analysis of apo D14 (PDB 3W04) and D14 bound to D3-CTH, which highlights a possible allosteric pathway that connects Leu707 of D3-CTH to the catalytic triad of D14. Arrows indicate conformational changes within D14 that are induced by binding to D3-CTH.
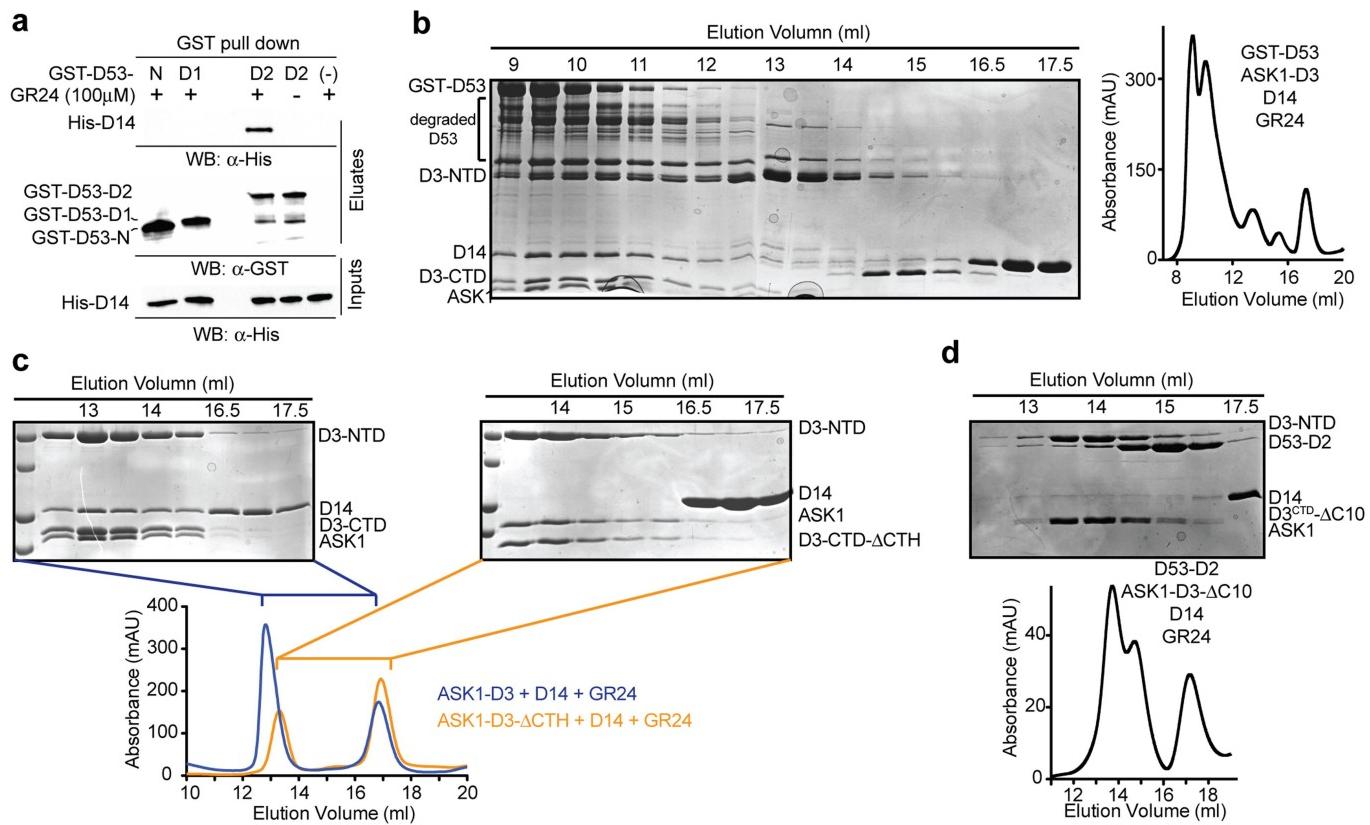
Extended Data Fig. 6 | The formation of the D3–D14–D53 complex is mediated by the D2 domain of D53. a, Pull-down assay using recombinant ASK1–D3, His–D14, and GST-tagged N domain (D53-N), D1 domain (D53-D1) or D2 domain of D53. **b–d**, Size-exclusion chromatography analyses of the interaction between: full-length GST–D53, D14–GR24 and ASK1–D3 (**b**), D14–GR24 and either ASK1–D3 or ASK1–D3(ΔCTH) (**c**), and D14–GR24 and D53-D2 with ASK1–D3(ΔCTH) (**d**). All gels were resolved by SDS–PAGE and analysed by western blot using anti-GST and anti-His antibodies (as indicated under **a**) or Coomassie blue staining (**b–d**). All experiments shown in **a–d** were repeated independently at least three times.
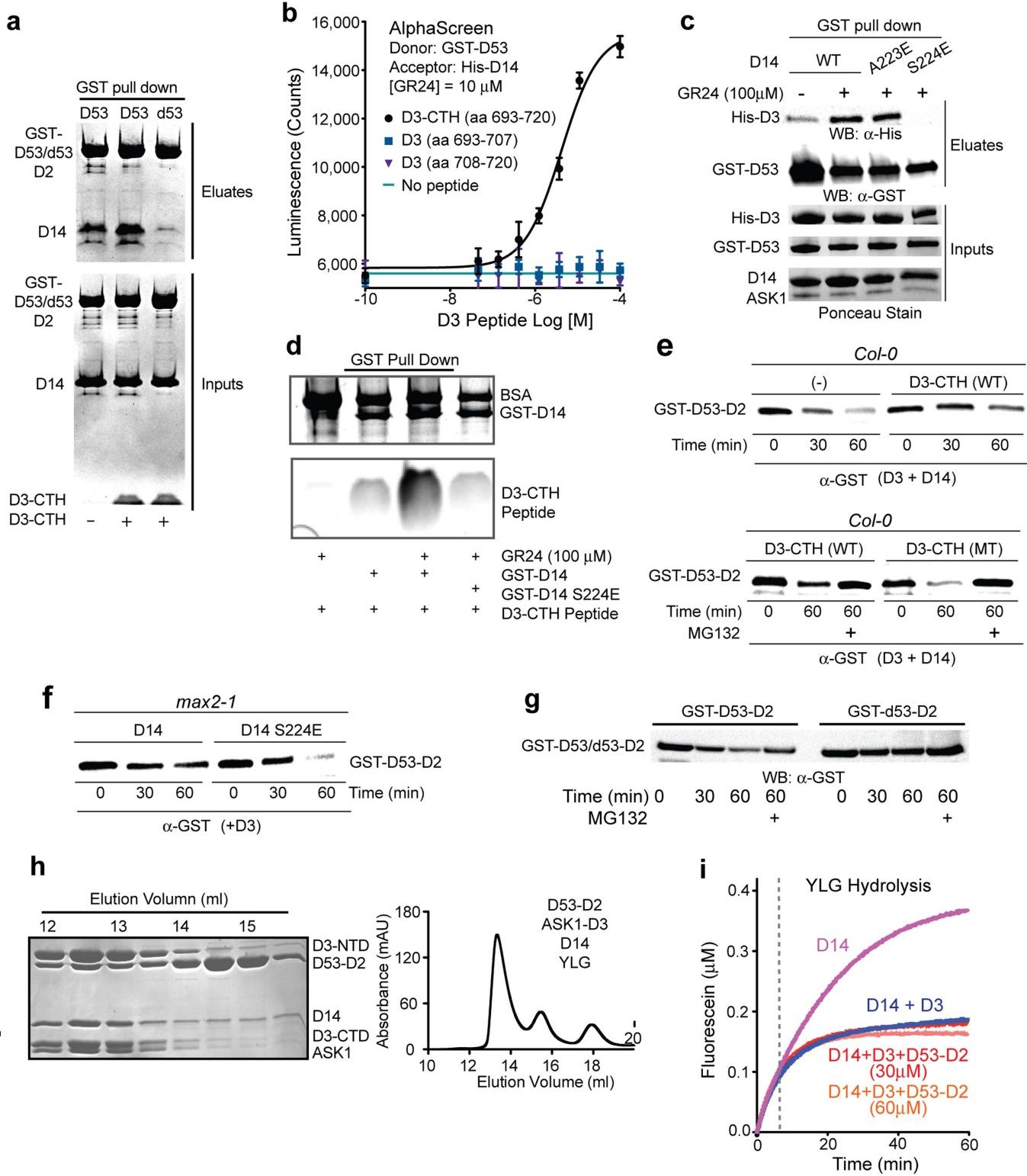
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | D3-CTH facilitates the binding of the D2 domain of D53 to D14–D3. a, GST pull-down assay using GST–D53-D2 or the GST-tagged D2 domain of the d53 (GST–d53-D2) mutant with non-tagged D14, in the presence or absence of D3-CTH as indicated. **b**, AlphaScreen data showing the ability of the D3-CTH peptide (28 amino acids, D3(693–720)) to promote the interaction between D53-D2 and D14 in a dose-dependent manner; D3(693–707) (15 amino acids) and D3(708–720) (13 amino acids) peptides did not stimulate binding. DMSO (indicated as 'no peptide') served as control (data are mean ± s.d. of biological triplicates). **c**, GST pull down using recombinant GST–D53-D2 and His–D3-ASK1 in the presence of recombinant D14 wild type (WT), D14(A223E), D14(S224E) and GR24 as indicated. **d**, GST pull down in the presence of the D3-CTH peptide with or without GR24, and in the presence of GST–D14 wild type or GST–D14(S224E). BSA was used in the assay to prevent non-specific interactions. MG132 was added as indicated. Proteins were resolved using SDS–PAGE, and were visualized by Coomassie blue staining or western-blot using anti-GST antibodies. The D3-CTH peptide contains four amino acid mutations that were designed to disrupt the D14–D3-CTH interface: E700R, L707R, D719R and D720R. **e**, **f**, Degradation of GST–D53-D2 in the *Col-0* (**e**) or *max2-1* (**f**) *A. thaliana* cell-free extract system. GST–D53-D2 was resolved at the indicated time in the presence or absence of the wild-type D3-CTH peptide (**e**, top) or a mutant (MT) (**e**, bottom), and in the presence of D3 and either D14 wild type or the D14(S224E) mutant (**f**). **g**, Time-dependent degradation of GST–D53-D2 and GST–d53-D2 in *Arabidopsis* seedlings of Col-0 extracts. Proteins were resolved by SDS–PAGE, and analysed by western blot using anti-GST antibody. MG132 indicates the addition of proteasome inhibitor. **h**, Size-exclusion chromatography analysis of complex formation among D53-D2, ASK1–D3 and D14 in the presence of YLG. **i**, Kinetics of YLG hydrolysis by D14 in the presence of ASK1–D3 and D53-D2 at two concentrations. Gels were resolved by SDS–PAGE and analysed by western blot using anti-GST and anti-His antibodies as indicated under **c**, **e**–**g**. All experiments were repeated independently at least three times.
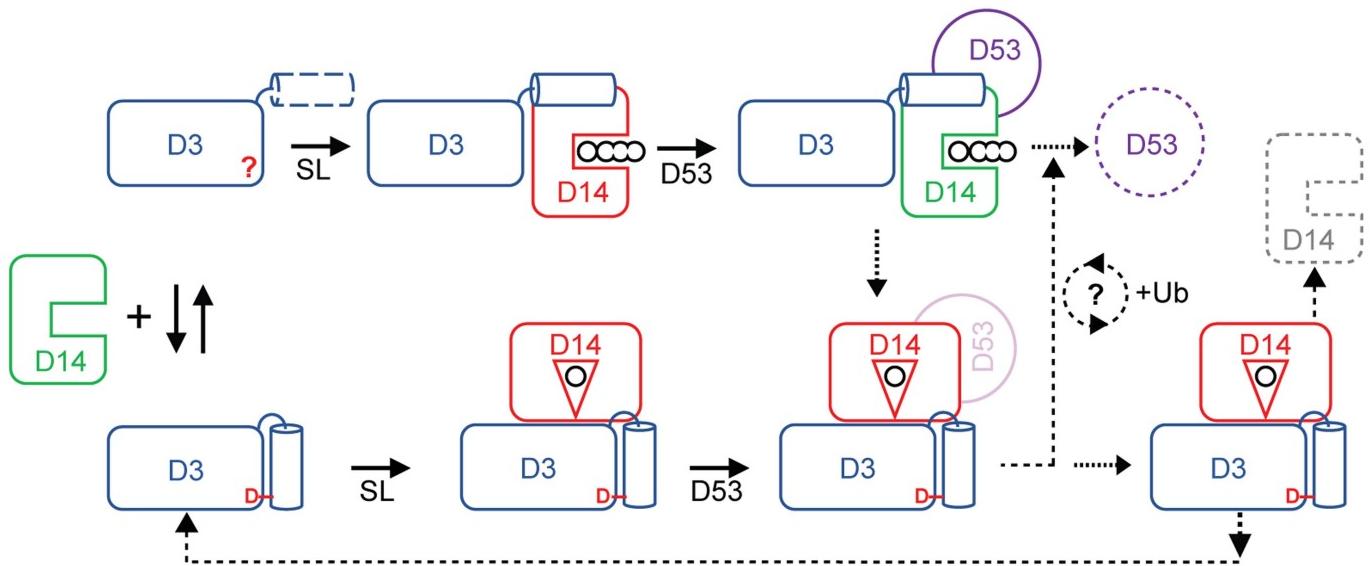
Extended Data Fig. 8 | A model for strigolactone perception and signalling. A model of the activity cycle that underlies strigolactone-induced and SCF^{D3-D14}-mediated D53 polyubiquitination. D3 adopts two conformational states with a structurally variable CTH (left). With a dislodged CTH, D3 binds and inhibits D14 in its open conformation, until D53 is loaded (top). D53 binding re-activates D14, which can hydrolyse strigolactones after or while D53 is polyubiquitinated. The strigolactone hydrolysis intermediate then stabilizes the closed conformation of D14, which converts D3-CTH into its engaged form. The resulting complex can ubiquitinate D14 and feed D3 back to the activity cycle (right). CLIM-bound D14 might participate in D53 polyubiquitination or in an alternative path (bottom). It remains unknown how many strigolactone molecules are hydrolysed during the polyubiquitination of each D53 molecule.

Extended Data Table 1 | Data collection and refinement statistics

	ASK1-D3 (form 1)		ASK1-D3 (form 2)	ASK1-D3 (form 3)	D3-CTH-D14-GR24
	Native	K$_2$Pt(NO$_2$)$_4$			
Data collection					
Space group	C2	C2	P2$_1$	P2$_1$	P6$_5$
Cell dimensions					
a, b, c (Å)	233.7 79.7 153.4	237.4 79.8 151.7	79.4 130.4 94.3	77.9 113.3 92.8	183.8 183.8 153.6
α, β, γ (°)	90 128.6 90	90 129.7 90	90 99.4 90	90 99 90	90 90 120
Resolution (Å)	50.00 - 2.50 (2.54 - 2.50)	50.00 - 2.50 (2.54 – 2.50)	50.00 - 2.40 (2.49-2.40)	50.00-3.00 (3.05-3.00)	50.00-2.40 (2.44-2.40)
R_{sym}	0.126 (0.403)	0.128 (0.539)	0.179 (0.939)	0.129(0.658)	0.172(0.699)
$I / \sigma I$	35.5 (2.0)	63.7 (2.0)	26.4 (3.0)	39 (1.3)	60.1 (2.5)
Completeness (%)	98.8 (88.4)	95.9 (71.6)	98.9 (97.8)	96.6 (97.8)	100 (99.8)
Redundancy	4.7 (3.6)	8.1 (6.3)	7.2 (5.6)	3.7 (3.5)	11.7 (8.9)
Refinement					
Resolution (Å)	2.50		2.40	3.00	2.40
No. reflections	75577		73464	31050	115753
R_{work} / R_{free} (%)	20.0/22.5		19.2/21.8	22.0/25.6	24.8/30.0
No. atoms	11832		12231	10771	17166
Protein	11366		11737	10771	16607
Ligand/ion	0		0	0	8
Water	466		494	0	551
B-factors					
Protein	39.3		35.5	65.6	32.4
Ligand/ion	N/A		N/A	N/A	38.6
Water	34.9		31.9	N/A	24.1
R.m.s. deviations					
Bond lengths (Å)	0.010		0.010	0.008	0.011
Bond angles (°)	1.48		1.40	1.31	1.27
Ramachandran favored (%)	97.9		98.5	96.5	96.0
Ramachandran allowed (%)	2.1		1.5	3.5	4.0
Ramachandran outliers (%)	0		0	0	0
PDB ID	6BRO		6BRP	6BRQ	6BRT

This table describes the data collection, phasing and refinement statistics of ASK1–D3 crystals in three forms, as well as crystals of D3-CTH–D14–GR24. Values in parentheses are for the highest-resolution shell.

LETTER

Spatially resolved rotation of the broad-line region of a quasar at sub-parsec scale

GRAVITY Collaboration*

The broadening of atomic emission lines by high-velocity motion of gas near accreting supermassive black holes is an observational hallmark of quasars[1]. Observations of broad emission lines could potentially constrain the mechanism for transporting gas inwards through accretion disks or outwards through winds[2]. The size of regions for which broad emission lines are observed (broad-line regions) has been estimated by measuring the delay in light travel time between the variable brightness of the accretion disk continuum and the emission lines[3]—a method known as reverberation mapping. In some models the emission lines arise from a continuous outflow[4], whereas in others they arise from orbiting gas clouds[5]. Directly imaging such regions has not hitherto been possible because of their small angular size (less than 10^{-4} arcseconds[3,6]). Here we report a spatial offset (with a spatial resolution of 10^{-5} arcseconds, or about 0.03 parsecs for a distance of 550 million parsecs) between the red and blue photo-centres of the broad Paschen-α line of the quasar 3C 273 perpendicular to the direction of its radio jet. This spatial offset corresponds to a gradient in the velocity of the gas and thus implies that the gas is orbiting the central supermassive black hole. The data are well fitted by a broad-line-region model of a thick disk of gravitationally bound material orbiting a black hole of 3×10^8 solar masses. We infer a disk radius of 150 light days; a radius of 100–400 light days was found previously using reverberation mapping[7–9]. The rotation axis of the disk aligns in inclination and position angle with the radio jet. Our results support the methods that are often used to estimate the masses of accreting supermassive black holes and to study their evolution over cosmic time.

We observed the quasar 3C 273 at the Very Large Telescope Interferometer (VLTI) in Chile using the recently deployed GRAVITY instrument[10] on eight nights between July 2017 and May 2018. The instrument coherently combines the light of the four 8-m telescopes to form interferometric amplitudes and phases on each of the six baselines (telescope pairs). The amplitudes measure the angular extent (size) of a structure, whereas the phases provide its on-sky position. The continuum dust emission was partially resolved (diameter of about 0.3 mas) and the broad-line region was more compact. We extracted differential phase curves (interferometric phase as a function of wavelength, measured relative to the continuum) for each of the six baselines near the (cosmologically) redshifted Paα line, which was observed at a wavelength of $\lambda \approx 2.17\,\mu$m, and averaged them over time to increase the signal-to-noise ratio. The differential phase ($\Delta\phi$) measures the shift in the photo-centre on the sky (Δx) of the total (line + continuum) image along the projected baseline direction for an unresolved source:

$$\Delta\phi(\lambda) = -\,2\pi f_{\mathrm{line}}\frac{B}{\lambda}\Delta x(\lambda) \qquad (1)$$

where B is the sky-projected length of the baseline (telescope separation) and f_{line} is the ratio of the emission line to the total flux in each wavelength channel (Methods). In this way, precise phase measurements from spectro-astrometry[11] provide spatial information on scales much smaller than the interferometric beam. 3C 273 is the most

attractive target for spectro-astrometry because it is the brightest nearby quasar, with a large region size measured from reverberation mapping[7–9], and its strong Paα line is observable in the near-infrared K band in which GRAVITY operates.

The strength of the phase signal depends on the kinematics. Turbulent motion produces zero differential phase, whereas a spatial velocity gradient results in wavelength-dependent phase shifts. We detected such a velocity gradient on three of our six baselines (UT4-3, UT4-2 and UT4-1; Methods)—those that were not aligned with the direction of the jet of 3C 273. Averaging the data of these three baselines, we find phase peaks of $0.25° \pm 0.06°$ over multiple spectral channels (Fig. 1a). We reject the null hypothesis of zero differential phase at a significance of more than 5σ (Methods). This precision in the differential phase is more than ten times better than that achieved previously in interferometry of active galactic nuclei (AGN)[12].

Using the differential phase data from all baselines, we measure the two-dimensional position of the photo-centre (the model-independent image centroid) in each wavelength channel that contributes sufficient line flux that equation (1) can be inverted ($f_{\mathrm{line}} > 0.35$; seven channels). We clearly detect a velocity gradient (Fig. 1b, spatial separation between red and blue sides of the line), which is nearly perpendicular to the large-scale position angle of the radio jet of 3C 273[13] (222°). By fitting a model with symmetric photo-centres on the red and blue sides of the line to all data, we determine a measured offset of $\Delta x = \pm(-9.5, 6.8) \pm (1.6, 1.1)\,\mu$as in right ascension and declination to the blue and red sides, which corresponds to a radial shift of about 0.03 pc (6,000 AU). This precision of about 1 μas corresponds to about 500 AU at a redshift of $z = 0.158$ (550 Mpc).

A velocity gradient perpendicular to the jet is strong evidence of 'ordered' rotation, although the signal-to-noise ratio in each channel is too low to uniquely determine the rotation profile. Here we use 'ordered' in the sense of coherent motion that produces a velocity gradient. For example, orbital motions with random angular momenta would not produce the phase signature that we detect.

To constrain the size and structure of the line-emitting region, we adopt a kinematic model of a collection of non-interacting point particles ('clouds') with equal (arbitrary) line flux distributed in an annular thick disk geometry and assuming circular orbits in the gravitational potential of the black hole[14,15] (Methods). The seven model parameters that we use are the mean and inner radii of the broad Paα line-emitting region (BLR) on the sky (R_{BLR} and R_{min}), its radial shape parameter (β, ranging from concentrated around the mean to exponentially distributed), the opening angle of the disk (θ_{o}), the inclination and position angles of the observer (i and PA), and the mass of the central black hole (M_{BH}; see Fig. 1c). We find a good joint fit (reduced $\chi^2 \approx 1.3$) to the Paα line profile and to the interferometric phases on all baselines and 40 wavelength channels, and constrain the model parameters using Markov chain Monte Carlo sampling[16] (Table 1, Methods). The centroids in each channel found by the model (Fig. 1b, dashed line) are in good agreement with the observed centroids.

We infer a position angle of PA $= 210°^{+6°}_{-9°}$ (with a 180° ambiguity) and an inclination of $i = 12° \pm 2°$ (all intervals at 90% confidence) for
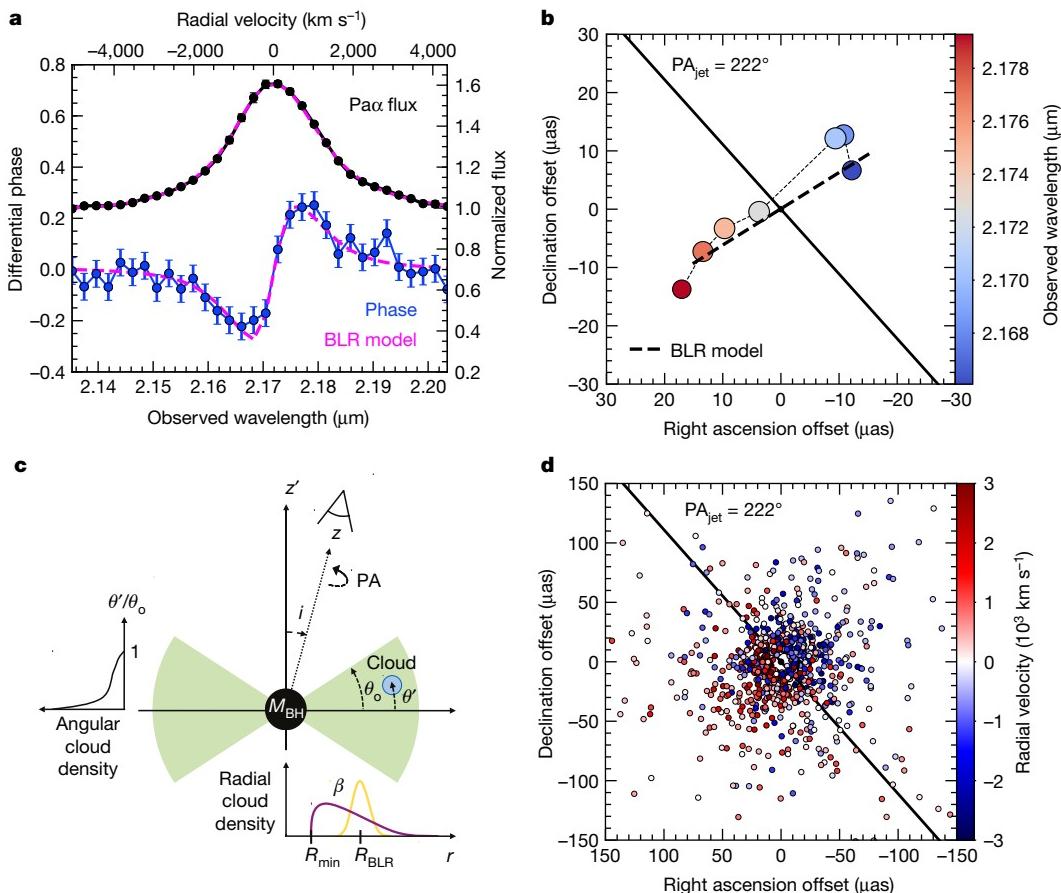
Fig. 1 | Main observational and modelling results. a, Paα line profile (black points; right axis) of 3C 273 observed by GRAVITY, along with the differential phase averaged over three baselines (blue points; left axis), showing the 'S' shape that is typical for a velocity gradient. The error bars represent 1σ. A thick-disk model of the BLR (dashed pink lines, see also **c** and **d**) provides an excellent joint fit to the data. **b**, The observed centroid position of the photo-centre in several wavelength channels (indicated by the colour scale; symbol size is proportional to the signal-to-noise ratio) show a clear spatial separation between redshifted and blueshifted emission: a velocity gradient at a position angle (PA) nearly perpendicular to that of the radio jet (PA$_{jet}$; solid black line). The centroid track of the BLR model is shown as a dashed line. **c**, Schematic representation of the model parameters. The green shaded area shows the geometry of the gas that surrounds the supermassive black hole (of mass M_{BH}), with the blue circle indicating an individual gas cloud. The angular (θ'; normalized by the opening angle of the disk θ_o) and radial (r) distributions of the gas clouds are plotted on the left and below, respectively. The rotation axis of the disk points along z', which is inclined by an angle i to the line of sight z. **d**, Velocity map (colour scale) of the model that best fits the discrete clouds (points)—a thick disk geometry viewed nearly face-on. Disorder in the velocity map reduces the observed shifts in the photo-centre (**b**) compared to the angular size of the BLR (**d**).

the rotation axis of the BLR. Measurements[17–19] of the inclination angle of the radio jet from superluminal motion range from 7° to 15°. The close two-dimensional alignment of the rotation axis and the radio jet confirms that the kinematics are dominated by ordered rotation. The half opening angle of the gas distribution is $45^{\circ+9^{\circ}}_{-6^{\circ}}$.

The mean radius of the Paα emitting region is found to be $R_{BLR} = 46 \pm 10$ μas (0.12 ± 0.03 pc), with an inner edge at $R_{min} = 11 \pm 3$ μas (0.03 ± 0.01 pc) and a roughly exponential radial emission profile, with shape parameter $\beta = 1.4 \pm 0.2$ (Methods). The measured mean radius corresponds to 145 ± 35 light days, roughly half the values obtained from previous reverberation mapping estimates (260–380 light days) using Hβ and Hγ emission lines[7,8], but consistent with the lower limit of 100 light days found from a subsequent re-analysis[9]. The discrepancy is probably due to the difficulty in measuring long lags in the brightest quasars, and could be partially due to intrinsic source variability (Methods).

Our inferred BLR radius is also a factor of roughly three smaller than the continuum dust radius found from previous interferometry measurements[20] and from our own data. This has been found to be the case for many AGN[21]. The size is also much smaller than that found tentatively from previous spectro-interferometry of 3C 273 using the VLTI instrument AMBER[12] ($R_{BLR} \approx 200$ μas). The object was too faint for fringe tracking in those observations. In addition, one of the three baselines used by AMBER was in the direction of the jet, that is, perpendicular to the disk. Our GRAVITY data, which have higher precision by a factor of about 40, rule out such a large size (Extended Data Fig. 4).

The inferred inner edge of the Paα emission region is a result of the cut-off in the line profile at $\pm 4{,}000$ km s^{-1}, which probably corresponds to the location where Balmer and Paschen emission becomes weak compared with that of higher-ionization lines[22]. The best-fitting structure is similar to that found from velocity-resolved reverberation mapping of nearby Seyfert 1 galaxies[23,24], which suggests that the properties of BLRs may not vary strongly with the luminosity or Eddington ratio of AGN.

We infer the black-hole mass of 3C 273 directly from the model to be $M_{BH} = (2.6 \pm 1.1) \times 10^8 M_\odot$, where $M_\odot$ is the solar mass. In reverberation mapping experiments, M_{BH} is obtained by combining Balmer-line time-delay measurements with the gas velocity obtained from the line profile. This requires the use of a velocity-inclination factor $f = GM_{BH}/(v^2 R_{BLR})$, where G is the gravitational constant and v is the gas velocity, which is usually defined as either the second moment of the line profile σ_{line} or the full-width at half-maximum of the line FWHM$_{line}$. Empirical mean values of f ($\langle f \rangle$) are obtained by matching the mass estimates from reverberation mapping with those from the relationship between the black-hole mass M_{BH} and the stellar velocity

Table 1 | Estimates of the kinematic BLR model parameters

Parameter	Value	Description
R_{BLR} (μas)	46 ± 10	Mean angular distance of the cloud from the black hole
R_{min} (μas)	11 ± 3	Minimum angular distance of the cloud from the black hole
β	1.4 ± 0.2	Radial distribution shape parameter
θ_o (°)	45^{+9}_{-6}	Half-opening angle of the disk
i (°)	12 ± 2	Inclination angle of the observer
PA (° E of N)	210^{+6}_{-9}	Position angle on the sky
M_{BH} ($10^8 M_\odot$)	2.6 ± 1.1	Black-hole mass

Values are medians and uncertainties represent 90% confidence ranges.

dispersion of galaxy bulges[25,26]. Typically, $\langle f \rangle = 4.3$ for $v = \sigma_{line}$ and $\langle f \rangle = 1.1$ for $v = \mathrm{FWHM}_{line}$. For the observed 3C 273 Paα line, we measure $\sigma_{line} = 1{,}400$ km s^{-1} ($f = 4.7 \pm 1.4$) and $\mathrm{FWHM}_{line} = 2{,}700$ km s^{-1} ($f = 1.3 \pm 0.2$), both in good agreement with the mean values used for Hβ in the literature[27].

Because they used a larger R_{BLR} (260–380 light days)[7,8,27] for 3C 273, previous studies found a mass of roughly double what we found. The mass is correlated with inclination: higher mass requires lower inclination to match the observed line width. Our R_{BLR} and M_{BH} could also in principle be underestimated: a larger BLR with added disorder may also be able to explain the phase signatures that we observe. Our highest-velocity photo-centre on the blue side (which has the lowest signal-to-noise ratio and is therefore not shown in Fig. 1) is offset (by 1σ–2σ) to the south and lies along the jet (Methods, Extended Data Fig. 2). This could be a hint of a second kinematic component contributing line flux.

Our results support the fundamental assumptions of reverberation mapping. The broad line width is dominated by bound motion in the gravitational potential of the black hole, and our inferred values of f match what is commonly found from the mean population. The size of the region is about half what is typically assumed, but within the large range of results reported for 3C 273. Quantitative size comparisons should be done using Seyfert galaxies with shorter lags, for which the radius is better determined from reverberation mapping. A comparison between velocity-resolved reverberation and (contemporaneous) interferometry in the same object would be particularly promising: reverberation mapping can unambiguously distinguish rotation from inflow and outflow, interferometry measures the orientation on the sky, and both techniques independently probe size and structure. The thick disk structure that we infer in 3C 273 could have a physical origin either in the inflowing surface layer of a geometrically thick accretion disk[28] or in a rotating wind launched from the surface of a thin disk[29]. Distinguishing these scenarios will require detailed model predictions, because strong outflows can still exhibit dominant signatures of rotation[30].

The combination of low inclination angles, large opening angles, a small radius and the alignment of our baselines with the rotation axis reduces the observed phase amplitude (roughly 0.3°) from the predicted value (more than 1°). Nonetheless, we have shown that the sensitivity of GRAVITY allows measurements of the size and structure of the BLR. Future observations could aim to measure a rotation curve and search for additional (for example, outflow) components and structural variability. With current instrumentation, spectro-interferometry is limited to the brightest AGN (with near-infrared magnitudes of less than 12 and visible magnitudes of less than 15). With an on-going observing programme, we plan to measure the properties of line and torus emission regions in a small sample of quasars and Seyfert galaxies[10]. The angular size of each component scales with the observed optical flux[6], making interferometry particularly well suited to exploring the physical origin of the broad-line region and to measuring black-hole masses in luminous quasars such as 3C 273, which are more representative of the supermassive black holes found in large samples out to high redshift.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0731-9.

1. Schmidt, M. 3C 273: a star-like object with large red-shift. *Nature* **197**, 1040 (1963).
2. Krolik, J. H. *Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment* (Princeton Univ. Press, Princeton, 2001).
3. Blandford, R. D. & McKee, C. F. Reverberation mapping of the emission line regions of Seyfert galaxies and quasars. *Astrophys. J.* **255**, 419–439 (1982).
4. Chiang, J. & Murray, N. Reverberation mapping and the disk-wind model of the broad-line region. *Astrophys. J.* **466**, 704–712 (1996).
5. Eracleous, M. & Halpern, J. P. Double-peaked emission lines in active galactic nuclei. *Astrophys. J. Suppl. Ser.* **90**, 1–30 (1994).
6. Elvis, M. & Karovska, M. Quasar parallax: a method for determining direct geometrical distances to quasars. *Astrophys. J.* **581**, L67–L70 (2002).
7. Kaspi, S. et al. Reverberation measurements for 17 quasars and the size-mass-luminosity relations in active galactic nuclei. *Astrophys. J.* **533**, 631–649 (2000).
8. Peterson, B. M. et al. Central masses and broad-line region sizes of active galactic nuclei. II. A homogeneous analysis of a large reverberation-mapping database. *Astrophys. J.* **613**, 682–699 (2004).
9. Zu, Y., Kochanek, C. S. & Peterson, B. M. An alternative approach to measuring reverberation lags in active galactic nuclei. *Astrophys. J.* **735**, 80 (2011).
10. GRAVITY Collaboration. First light for GRAVITY: phase referencing optical interferometry for the Very Large Telescope Interferometer. *Astron. Astrophys.* **602**, A94 (2017).
11. Bailey, J. Detection of pre-main-sequence binaries using spectro-astrometry. *Mon. Not. R. Astron. Soc.* **301**, 161–167 (1998).
12. Petrov, R. G. et al. VLTI/AMBER differential interferometry of the broad-line region of the quasar 3C273. *Proc. SPIE* **8445**, 84450W (2012).
13. Röser, H.-J. & Meisenheimer, K. The synchrotron light from the jet of 3C 273. *Astron. Astrophys.* **252**, 458–474 (1991).
14. Pancoast, A., Brewer, B. J. & Treu, T. Modelling reverberation mapping data – I. Improved geometric and dynamical models and comparison with cross-correlation results. *Mon. Not. R. Astron. Soc.* **445**, 3055–3072 (2014).
15. Rakshit, S., Petrov, R. G., Meilland, A. & Hönig, S. F. Differential interferometry of QSO broad-line region – I. Improving the reverberation mapping model fits and black hole mass estimates. *Mon. Not. R. Astron. Soc.* **447**, 2420–2436 (2015).
16. Foreman-Mackey, D. emcee: the MCMC hammer. *Publ. Astron. Soc. Pacif.* **125**, 306–312 (2013).
17. Lobanov, A. et al. A cosmic double helix in the archetypical quasar 3C273. *Science* **294**, 128–131 (2001).
18. Savolainen, T. et al. Multifrequency VLBA monitoring of 3C 273 during the INTEGRAL Campaign in 2003 I. Kinematics of the parsec scale jet from 43 GHz data. *Astron. Astrophys.* **446**, 71–85 (2006).
19. Jorstad, S. G. et al. Kinematics of parsec-scale jets of gamma-ray blazars at 43 GHz within the VLBA-BU-BLAZAR program. *Astrophys. J.* **846**, 98 (2017).
20. Kishimoto, M. et al. The innermost dusty structure in active galactic nuclei as probed by the Keck interferometer. *Astron. Astrophys.* **527**, A121 (2011).
21. Netzer, H. Revisiting the unified model of active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **53**, 365–408 (2015).
22. Baldwin, J. A. Luminosity indicators in the spectra of quasi-stellar objects. *Astrophys. J.* **214**, 679–684 (1977).
23. Pancoast, A. et al. Modelling reverberation mapping data – II. Dynamical modelling of the Lick AGN monitoring project 2008 data set. *Mon. Not. R. Astron. Soc.* **445**, 3073–3091 (2014).
24. Grier, A. et al. The structure of the broad-line region in active galactic nuclei. II. Dynamical modeling of data from the AGN10 reverberation mapping campaign. *Astrophys. J.* **849**, 146 (2017).
25. Ferrarese, L. & Merritt, D. A fundamental relation between supermassive black holes and their host galaxies. *Astrophys. J.* **539**, L9–L12 (2000).
26. Gebhardt, K. et al. A relationship between nuclear black hole mass and galaxy velocity dispersion. *Astrophys. J.* **539**, L13–L16 (2000).
27. Bentz, M. C. & Sarah, K. The AGN black hole mass database. *Publ. Astron. Soc. Pacif.* **127**, 67–73 (2015).
28. Begelman, M. C. & Silk, J. Magnetically elevated accretion discs in active galactic nuclei: broad emission-line regions and associated star formation. *Mon. Not. R. Astron. Soc.* **464**, 2311–2317 (2017).
29. Czerny, B. et al. Failed radiatively accelerated dusty outflow model of the broad line region in active galactic nuclei. I. Analytical solution. *Astrophys. J.* **846**, 154 (2017).
30. Mangham, S. W. et al. The reverberation signatures of rotating disc winds in active galactic nuclei. *Mon. Not. R. Astron. Soc.* **471**, 4788–4801 (2017).

GRAVITY Collaboration
E. Sturm[1,19]*, J. Dexter[1,19]*, O. Pfuhl[1], M. R. Stock[1], R. I. Davies[1], D. Lutz[1], Y. Clénet[2], A. Eckart[3,4], F. Eisenhauer[1], R. Genzel[1,5], D. Gratadour[2], S. F. Hönig[6], M. Kishimoto[7], S. Lacour[2], F. Millour[8], H. Netzer[9], G. Perrin[2], B. M. Peterson[10,11,12], P. O. Petrucci[13], D. Rouan[2], I. Waisberg[1], J. Woillez[14], A. Amorim[15], W. Brandner[16], N. M. Förster Schreiber[1], P. J. V. Garcia[17,18], S. Gillessen[1], T. Ott[1], T. Paumard[2], K. Perraut[13], S. Scheithauer[16], C. Straubmeier[3], L. J. Tacconi[1] & F. Widmann[1]

[1]Max Planck Institute for Extraterrestrial Physics, Garching, Germany. [2]LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université Paris Diderot, Sorbonne Paris Cité, Meudon, France. [3]I. Physikalisches Institut, Universität zu Köln, Köln, Germany. [4]Max Planck Institute for Radio Astronomy, Bonn, Germany. [5]Departments of Physics and Astronomy, Le Conte Hall, University of California, Berkeley, CA, USA. [6]Department of Physics and Astronomy, University of Southampton, Southampton, UK. [7]Department of Physics, Kyoto Sangyo University, Kita-ku, Japan. [8]Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Nice, France. [9]School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel. [10]Department of Astronomy, The Ohio State University, Columbus, OH, USA. [11]Center for Cosmology and AstroParticle Physics, The Ohio State University, Columbus, OH, USA. [12]Space Telescope Science Institute, Baltimore, MD, USA. [13]Université Grenoble Alpes, CNRS, IPAG, Grenoble, France. [14]European Southern Observatory, Garching, Germany. [15]CENTRA, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal. [16]Max Planck Institute for Astronomy, Heidelberg, Germany. [17]CENTRA, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal. [18]European Southern Observatory, Vitacura, Chile. [19]These authors contributed equally: E. Sturm, J. Dexter. *e-mail: sturm@mpe.mpg.de; jdexter@mpe.mpg.de

METHODS

Observations. The observations were taken at the VLTI in Chile using the second-generation GRAVITY instrument[10] and the four 8-m unit telescopes. We chose the medium-resolution mode of GRAVITY with 90 independent spectral elements ($\lambda/\Delta\lambda \approx 500$) dispersed across 210 pixels. All data were obtained in single-field on-axis, combined polarization mode.

Each observation followed the same sequence. Each unit telescope locked its adaptive optics (MACAO) module on 3C 273. Once the adaptive optics loops were closed, the telescope beams were coarsely aligned on the VLTI laboratory camera IRIS. In a second step, the internal beam tracking of GRAVITY aligned the fringe-tracking and science fibres on the target.

The science exposures contain ten 30-s integrations (NDIT = 10 and DIT = 30 s). Fringe tracking on faint sources such as 3C 273 ($K \approx 10$, $V \approx 13$) is difficult, so we integrated deeply on-source, taking only occasional sky exposures, often at the end of the integration. In July 2017, January 2018 and May 2018, interferometric calibrators were also observed (although not required for differential phase measurements). Extended Data Table 1 lists the observing nights, the integration time, the seeing (as provided by the DIMM seeing monitor) and the coherence time.

Data reduction. We used the standard GRAVITY pipeline to process the data[10,31]. Each exposure consists of ten science frames, which are averaged after processing. Each frame is corrected for background bias by subtracting the closest sky exposure. The detector-noise bias is calculated from dark exposures with the same settings as the science exposures. Flat-field and wavelength calibration were done on the internal calibration source. The effective wavelength for each spectral element was calibrated by modulating the fibre lengths of the fringe-tracking and science channels and using the internal laser metrology as a wavelength reference.

The science data were reduced using a pixel-to-visibility matrix[32]. This matrix represents the matrix-encoded instrument transfer function, which includes the relative throughput, coherence, phase shift and cross-talk for each pixel. Applying the matrix to the detector frames yields the instrument-calibrated complex visibilities. In the next step, the complex visibilities from the science channel were phase-referenced to those from the fringe-tracking channel using the laser metrology. The effective optical path difference for each spectral channel was calculated from the delay measured by the laser metrology and the differential dispersion between the wavelengths of the laser and of the channel. This calculation yields phase-referenced complex visibilities. The GRAVITY pipeline removes a mean and slope from the raw visibility phase calculated using all wavelength channels to create a differential phase on each baseline.

We used an alternative method (developed for VLTI/AMBER[33]) and computed the differential phase of each channel as in equation (7) in ref. [33], which ignores the work channel when calculating the mean and slope. This method produces consistent results but improves our phase errors, typically by about 10%–20%. To account for the observatory transfer function (coherence loss due to vibrations, uncorrected atmosphere, birefringence, and so on), it is common practice to observe a calibrator star close to the science target. We did this for three epochs to obtain continuum visibilities.

Data processing. For each exposure, we further removed broad instrumental shapes in the differential phases. This was done by convolving the differential phase versus the wavelength from each baseline with a Gaussian with a FWHM of 24 pixels, and then dividing the actual data by this smoothed version. The width was chosen to flatten the differential phases without affecting their behaviour near the Paα line. After flattening, no calibrator shows any significant differential phase signature in any part of the spectrum, down to limits of less than 0.1°. Therefore, no further instrumental bias needs to be removed from the 3C 273 data.

From each night, we selected exposures for which (i) the fringe tracking was working more than 80% of the time, and (ii) the root-mean-square of the differential phase curves was less than 3° on all baselines. This selection removed 25% (26/104) of the 5-min exposures. Table 1 summarizes the data used in the analysis presented here.

We then averaged the exposures in time for each night for each of the four epochs to reduce the phase noise. For the first three epochs, the changes in the coordinates (u, v) of the visibility domain for each baseline are small during integrations of less than 1 h. In May 2018, the integrations were longer. We checked for differences between the first and second half of each night, and for the expected changes in phase signatures in the best-fitting model (see below). In both cases, the changes are smaller than our errors and so we consider it safe to time-average per baseline at all epochs. The weights for each exposure and baseline were calculated from the root-mean-square of the phase noise over a broad spectral region centred on the Paα line (2.06–2.28 μm). The averaged differential phase curves have a residual scatter of 0.2°–0.3° in each epoch. Individual 5-min exposures typically reached 0.7°–1.0°. In each epoch, the precision reached is a factor of roughly ten higher than previous spectro-interferometry of AGN[12].

We extracted the spectral line profile in each exposure by summing the photometric flux of each of the four telescopes. We removed the shape of the instrument response by dividing the 3C 273 spectrum by that of a calibrator star. Spectral lines from the calibrator star were removed by first dividing its spectrum by templates from the NASA Infrared Telescope database[34]. For March and April 2018, for which no calibrator was taken, we used that from the January 2018 observation instead. Then the red continuum slope was removed before averaging the result over all exposures within each epoch. The line shape and strength were stable between epochs to within 5%. That scatter could be due to either intrinsic variability or systematic error in the spectral extraction. Conservatively, we assumed that the line profile and flux are constant and averaged the four epochs to get a single line profile. The errors are taken as the quadrature sum of the root-mean-square between epochs and the statistical errors on the measured photometric flux.

Detection of a velocity gradient. As described in the main text, we first used the differential phase data to fit for a best centroid position at each spectral channel across the line. The model has an (x, y) position with a phase given by $\Delta\phi_{ij} = -2\pi f_{\text{line}}(u_j x_i + v_j y_i)$, where the index i corresponds to the seven spectral channels used (in which the line intensity relative to the continuum $f_i > 0.35$ with $f_{\text{line}} = f_i/(1 + f_i)$) and j corresponds to each of the 24 baselines (6×4 epochs). This is the form in the marginally resolved limit, in which we can expand the exponential in the complex visibility and keep only the first-order term[35,36]. In each channel we minimize the fit to the observed phases to find a best centroid position x_i. The results are shown in Extended Data Fig. 2, in which the red and blue wavelength channels clearly cluster on opposite sides of the line. The 1σ confidence intervals are shown as ellipses.

To estimate the significance of the detection, we consider a null hypothesis of zero phase everywhere and compare this to a model with single centroid positions (x, y) and $(-x, -y)$ in right ascension and declination on the blue and red sides of the line. We calculate the model phases as above for channels whose centroids appear to deviate significantly from zero (four red and four blue channels) and assign them zero phase elsewhere. From least-squares fitting, we find $(x, y) = (-9.5, 6.8) \pm (1.1, 1.6)$ μas. The χ^2 for the null hypothesis and model are 1,417 and 1,308, respectively, with 960 data points. An F-test rejects the $x = y = 0$ null hypothesis with a P value of 10^{-17}, corresponding to more than 8σ. The spectral channels are more finely sampled than their resolution element, so neighbouring channels are correlated. To estimate the effect of this correlation, we repeat the above test using half of the channels and find a P value for the null hypothesis of less than 10^{-8}, or a detection significance of more than 5.5σ.

The bluest channel in Extended Data Fig. 2 moves to the south, along the jet. Its phase agrees with those of other blue channels for the three off-jet baselines (Fig. 1a). However, it shows a $-0.2°$ phase in the average of the other three baselines. The significance is only 1σ–2σ, and we do not interpret it further here. If detected in future observations, the alignment with the large-scale jet direction could be a signature of outflowing gas at high velocity.

At certain baselines and spectral channels we see apparently systematic features that can extend over two to three channels, away from any spectral lines. None of these reaches close to the signal-to-noise ratio of the signature studied here. This is true for all GRAVITY AGN and calibrator data examined.

BLR model description and fitting procedure. We adopt a phenomenological model of the Paα emitting region to interpret our data, closely following previous work[14,15]. The model comprises a large number of non-interacting test particles on circular orbits around the central black hole under the sole influence of gravity. The particles represent dense, low-filling-factor, line-emitting gas clouds. Their distances from the black hole are

$$r = R_{\text{S}} + F R_{\text{BLR}} + g(1 - F)\beta^2 R_{\text{BLR}}$$

where $R_{\text{S}} = 2GM_{\text{BH}}/c^2$ is the Schwarzschild radius, R_{BLR} is the mean radius, $F = R_{\text{min}}/R_{\text{BLR}}$ is the fractional inner radius, β is the shape parameter and $g = p(x \mid 1/\beta^2, 1)$ is drawn randomly from a Gamma distribution:

$$p(x|\alpha, \theta) = \frac{x^{\alpha - 1}\exp(-x/\theta)}{\theta^\alpha \Gamma(\alpha)}$$

No line-emitting clouds are present inside R_{min} ($R_{\text{min}} \gg R_{\text{S}}$ for the viable models found for 3C 273), and their distribution is allowed to vary from a Gaussian distribution concentrated around the mean ($0 < \beta < 1$) to an exponential ($\beta = 1$) or steeper ($1 < \beta < 2$) profile concentrated at the inner radius. The angular distribution is specified by a half-thickness θ_{o}, and the clouds are placed at random positions along their orbits. The structure is viewed at an inclination angle i and rotated in the sky plane by a position angle PA measured in degrees east of north. In total, the model as implemented has seven free parameters: R_{BLR}, F, β, M_{BH}, θ_{o}, i and PA.

Recent velocity-resolved reverberation mapping studies used additional parameters describing inflow and outflow, anisotropic emission and asymmetry in the angular distribution[23,24]. Here we choose the minimal model required by our data.

Rotation explains the velocity gradient perpendicular to the jet axis, an inner radius is justified by cut-offs in the spectrum at roughly ±4,000 km s^{-1}, and steep radial and thick angular distributions can produce the observed line shape. Although we find a satisfactory fit (see below), including elliptical or radial motion would probably change our inferred parameters. In particular, additional model components that add disorder could potentially allow models with larger R_{BLR} and M_{BH} to fit the data.

Assuming that emitted Paα photons free-stream through the BLR once emitted, we conveniently obtain spectro-interferometric observables as sums over clouds binned by the observed wavelength λ_{obs}:

$$\frac{\lambda_{obs}}{\lambda_{emit}} = \left(1 + \frac{v_z}{c}\right)\left(1 - \frac{v^2}{c^2}\right)^{-1/2}\left(1 - \frac{R_S}{r}\right)^{-1/2} \quad (2)$$

with an emitted wavelength λ_{emit}, total velocity v and line-of-sight velocity v_z. We have included the relativistic and transverse Doppler shift and the gravitational redshift, because they could affect the line shape of the emission[37]. We note that equation (18) of ref. [14] neglects the transverse Doppler shift, which is of the same order as the gravitational redshift for orbital motion. We also note that in equation (2), $v_z < 0$ corresponds to motion towards the observer (blueshift), in keeping with the standard radial-velocity convention.

The spectral line shape is then found by summing the clouds in bins of observed wavelength. We account for the shape of the GRAVITY line spread function in medium resolution by binning at higher spectral resolution before convolving with a Gaussian with a FWHM of 4 nm. We then normalize the line strength so that it matches the observed strength at the peak and shift it to the observed central wavelength. The scaling and shift are both fixed in the analysis.

We model the differential phase of the BLR relative to a continuum that is assumed to be symmetric, as implied by observed closure phases in 3C 273 being less than 1°. Because the size of the region (less than 100 μas) is much less than the VLTI imaging resolution (about 3 mas), each cloud contributes a phase of $\phi_k(\lambda, u, v) = -2\pi f(\lambda)/[1 - f(\lambda)](ux_k + vy_k)$ in radians, where $f(\lambda)$ is the line intensity relative to the continuum and (u, v) is the baseline separation. The Fourier transform is linear, so we can find the total phase for each baseline by summing the individual phases in wavelength bins, using the same procedure as above to account for the instrument spectral point spread function.

We fit the seven-parameter model to all observed 3C 273 spectral line and phase curves simultaneously (40 wavelength channels for the time-averaged spectrum and 24 baselines; data shown in Extended Data Fig. 1). The number of wavelength channels is chosen to fully cover the line profile as well as a small off-region, because the inner radius that we infer depends on the maximum observed radial velocity in the tails of the line. We use 2×10^5 clouds in the model, the minimum number so that the likelihood does not vary significantly between random instantiations. We use Bayesian statistics to measure confidence intervals on the model parameters. The priors are uniform over the following intervals: $\log(R_{BLR}/\mu as) \in (0, 4)$, $F \in (0, 1)$, $\beta \in (0, 2)$, $\log(M_{BH}/M_\odot) \in (6, 10)$, $\theta_o \in (0°, 90°)$, $i \in (0°, 90°)$ and $PA \in (0°, 360°)$. The posterior is sampled using the emcee Markov chain Monte Carlo code[16], with a joint likelihood including both the spectral and differential phase data across all spectral channels.

The sampling is run with 1,000 'walkers' independently exploring the parameter space, and converges within about 800 trials (a total of 8×10^5 samples of the likelihood). We checked the convergence by comparing confidence intervals taken from well-separated trials and by starting the fit from several initial guesses far from the final distribution. We report 90% confidence intervals as the 5th and 95th percentiles of samples in the one-dimensional distribution over each parameter. A corner plot of the full set of two- and one-dimensional distributions is shown in Extended Data Fig. 3.

Inferred BLR size compared with reverberation mapping. Previous reverberation mapping studies using Hβ and Hγ have found time lags of 260–380 light days[7,8], or a lower limit of 100 light days[9]. Our measured value of 145 ± 35 light days using Paα is within this range, but roughly half what is typically assumed. The wide range in the reported values is due to the difficulty of reverberation mapping measurements for luminous quasars that are equatorial on the sky, such as 3C 273. The light curves must be combined across long seasonal observing gaps, leading to aliasing and other problems with irregular sampling. We believe that this is the main contributor to the discrepancy. In addition, about 25% of this difference can be explained by the half-a-magnitude higher source luminosity during the previous

observations. Other contributing factors could be a real difference between the Paα and Balmer line-emission regions, but this seems unlikely because these lines share an upper atomic level. Our measured line width is slightly smaller than what has been measured for the other lines, implying a larger radius, but this could also be the result of complex kinematics that were not accounted for.

As discussed in the main text, our measured size is consistent with the other known properties of 3C 273. It is a factor of about three smaller than the hot dust continuum, similar to what is seen in other sources[21]. However, it is well below the prediction from the radius–luminosity relation based on Hβ (260–380 light days)[38]. Other recent reverberation mapping work on samples of more luminous sources find smaller source sizes and a flatter radius–luminosity relation[39,40], which would agree with our size estimate. Additional interferometry measurements, well suited to luminous quasars, could help to study this further.

Limits on radial inflow and outflow. With only detections of differential phases, spectro-interferometry cannot distinguish between rotational and ordered radial motion (inflow or outflow)[15]. Here we see that the velocity gradient is nearly perpendicular to the large-scale jet, and conclude that it indicates rotation. In a mixed model in which rotation and inflow/outflow are both present, the position angle of the velocity gradient changes depending on their relative strength. As a simple estimate of the allowed fraction of inflow and outflow, we assume that the true rotation axis aligns with the 3C 273 jet. The observed offset between the two would then be due to outflow, which moves the net velocity gradient towards the jet direction. Fitting our previous BLR model with a fixed position angle constrains the fraction of radial outflow at the local escape speed to be at most about 25%. This limit is qualitative and, for example, changes depending on the true position angle of the rotation axis and the implementation of the outflow model. In particular, disk-wind models are often rotation-dominated at low velocity[30] and would probably not be subject to this constraint.

Size comparison with previous 3C 273 near-infrared spectro-interferometry. Tentative evidence has been found[12] for a very large BLR ($R_{BLR} \approx 200$ μas) using VLTI/AMBER. This result came from differential visibility amplitude data, in which a claimed decrease in the wavelength-dependent change in the interferometric amplitude implies a larger image size of the line than of the continuum[35,36]. We show the differential amplitude data from our GRAVITY observations in Extended Data Fig. 4, averaged over our two longest baselines (UT4-1 and UT3-1) and over all epochs. At ten-times-better precision, we see no sign of the claimed decrease[12] at the line. Instead, we see a clear increase in the amplitude at the line, showing that it is more compact than the hot dust continuum emission[20], $R_{BLR} \ll 150$ μas. This compact size agrees with the small R_{BLR} that we inferred independently from modelling the Paα spectral line shape and the differential visibility phases as described above.

Data availability

The data were obtained at the VLTI of the European Southern Observatory (ESO), Paranal, Chile, and are available on the ESO archive (http://archive.eso.org/eso/eso_archive_main.html) under programme IDs 099.B-0606, 0100.B-0582 and 0101.B-0255.

31. Lapeyrere, V. et al. GRAVITY data reduction software. *Proc. SPIE* **9146**, 91462D (2014).
32. Tatulli, E. et al. Interferometric data reduction with AMBER/VLTI. Principle, estimators, and illustration. *Astron. Astrophys.* **464**, 29–42 (2007).
33. Millour, F. et al. AMBER closure and differential phases: accuracy and calibration with a beam commutation, optical and infrared interferometry. *Proc. SPIE* **7013**, 70131G (2008).
34. Rayner, J. T., Cushing, M. C. & Vacca, W. D. The Infrared Telescope Facility (IRTF) spectral library: cool stars. *Astrophys. J.* **185**, 289–432 (2009).
35. Lachaume, R. On marginally resolved objects in optical interferometry. *Astron. Astrophys.* **400**, 795–803 (2003).
36. Waisberg, I. et al. Submilliarcsecond optical interferometry of the high-mass X-ray binary BP Cru with VLTI/GRAVITY. *Astrophys. J.* **844**, 72 (2017).
37. Goad, M. R., Korista, K. T. & Ruff, A. J. The broad emission-line region: the confluence of the outer accretion disc with the inner edge of the dusty torus. *Mon. Not. R. Astron. Soc.* **426**, 3086–3111 (2012).
38. Bentz, M. C. et al. The low-luminosity end of the radius-luminosity relationship for active galactic nuclei. *Astrophys. J.* **767**, 149 (2013).
39. Du, P. et al. Supermassive black holes with high accretion rates in active galactic nuclei IX. 10 new observations of reverberation mapping and shortened Hβ lags. *Astrophys. J.* **856**, 6 (2018).
40. Grier, C. J. et al. The Sloan Digital Sky Survey reverberation mapping project: Hα and Hβ reverberation measurements from first-year spectroscopy and photometry. *Astrophys. J.* **851**, 21 (2017).
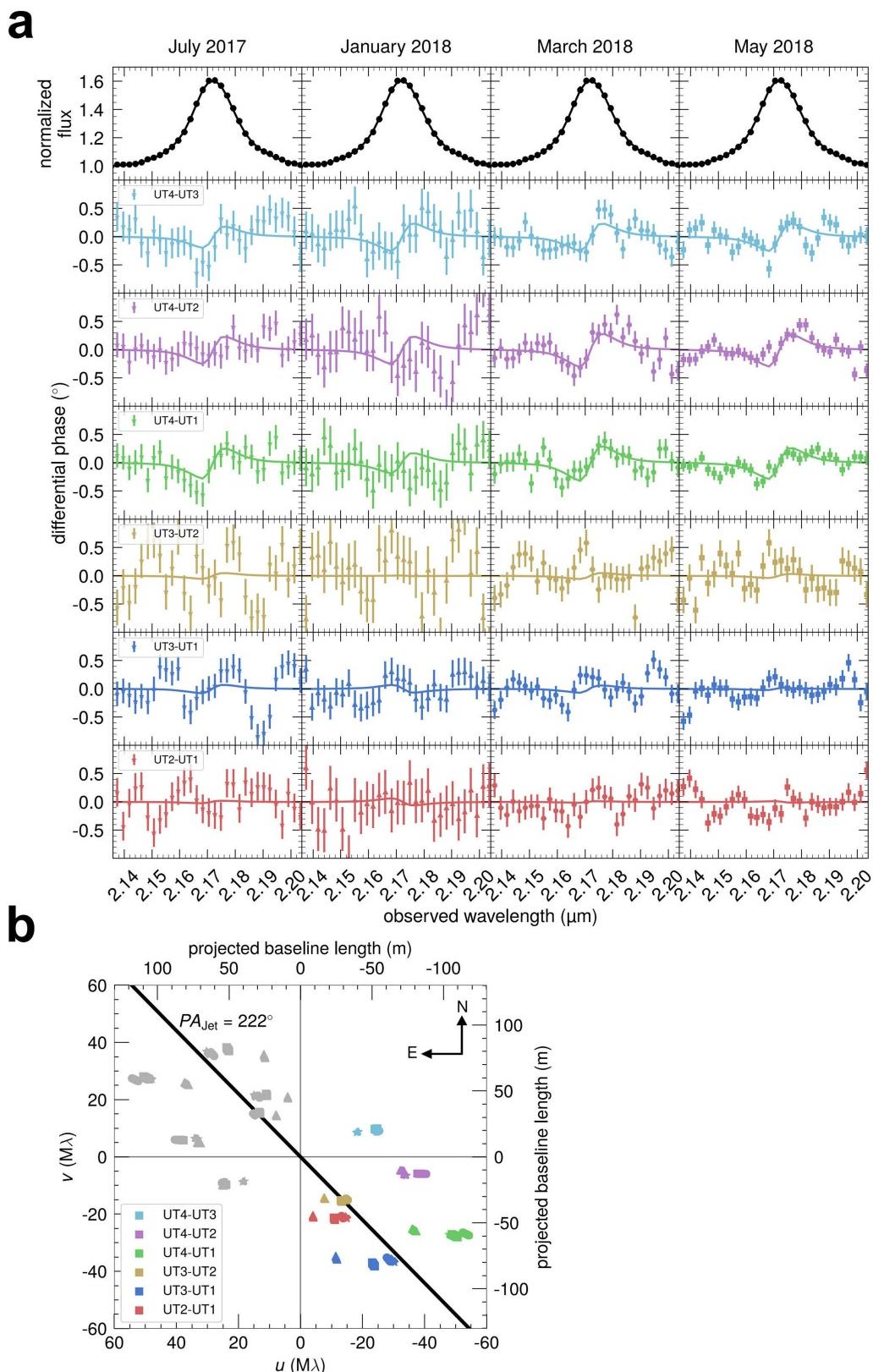
Extended Data Fig. 1 | Differential phases and u–v coverage.
a, Differential phase curves (coloured points with 1σ error bars) on the six VLTI baselines (rows) at four epochs (columns) and time-averaged 3C 273 Paα line profile (black points, identical in each panel). The best-fitting BLR model to all data are shown as solid lines. **b**, u–v coverage for all of our data (in units of millions of λ, Mλ), with observed points (in colour) mirrored in grey. Note the close alignment between all baselines, particularly those without UT4, with the position angle of the large-scale radio jet of 3C 273.
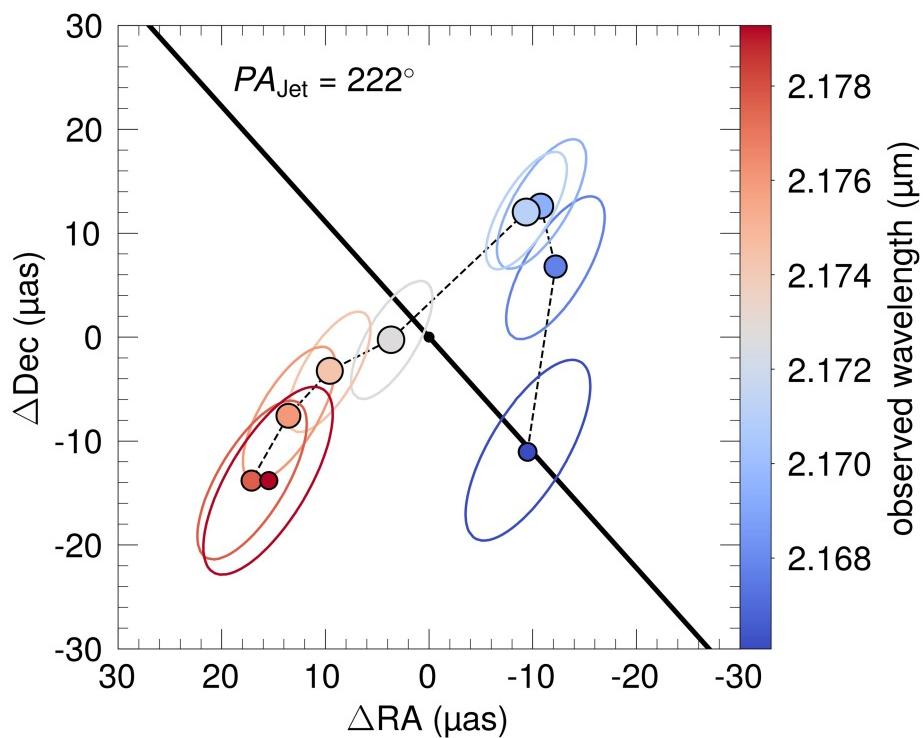
Extended Data Fig. 2 | Observed centroid positions in several wavelength channels. Best-fitting centroids to the differential phase data in each wavelength channel are shown as in Fig. 1, but with contour ellipses containing 68% of the probability density. In addition, the extremal points to the blue (on the jet axis) and red are not shown in Fig. 1, because of their larger errors. Given the relatively low signal-to-noise ratio in each channel, we cannot measure the radial velocity as a function of position (for example, the rotation curve). ΔDec, declination offset; ΔRA, right ascension offset.
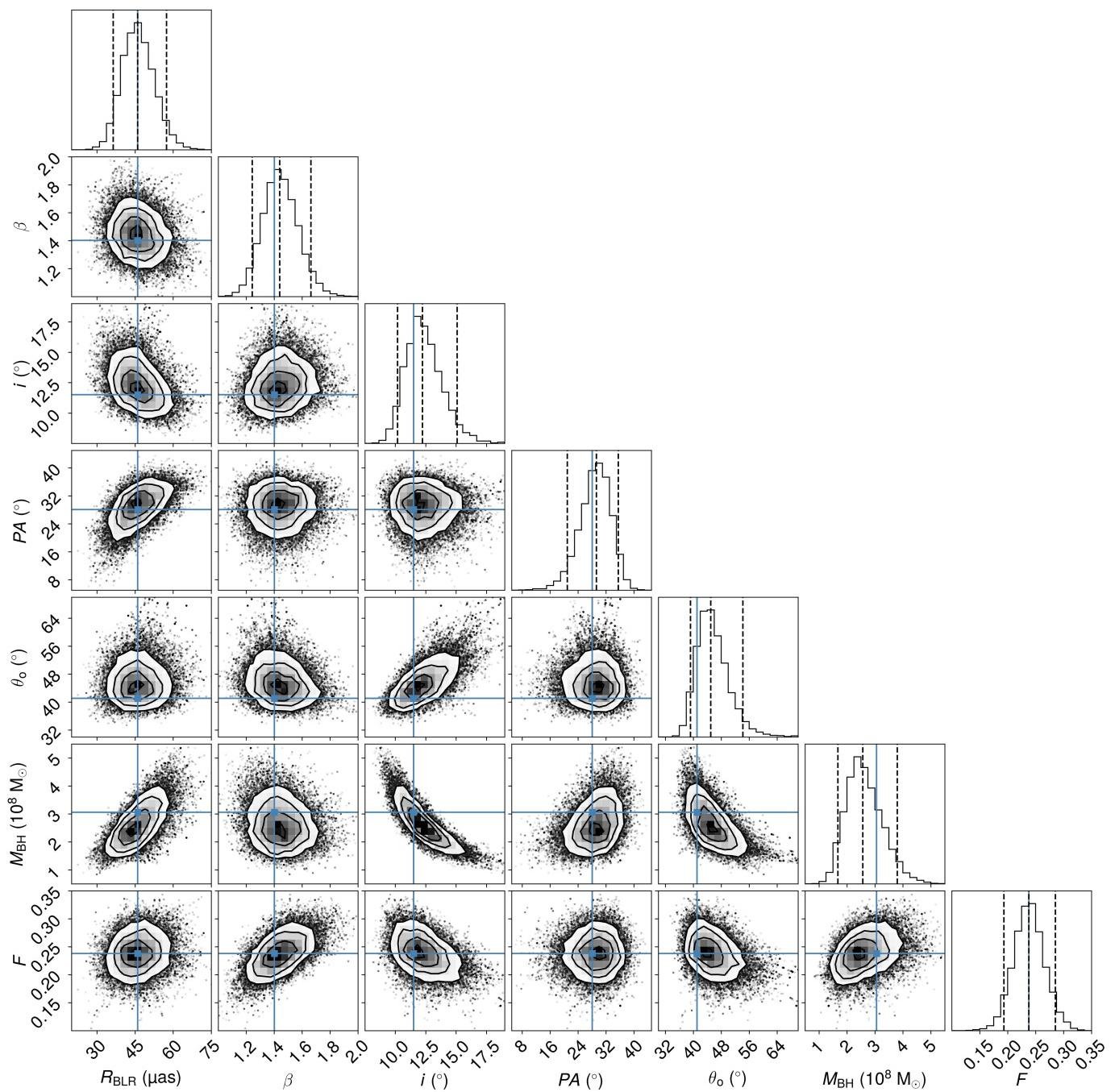
Extended Data Fig. 3 | Corner plot of the BLR model parameters.
One- and two-dimensional probability density distributions from fitting
the seven-parameter BLR model to the spectral line profile and differential
phase data for 3C 273 obtained from GRAVITY. The blue points show the
parameter values at which the highest likelihood was obtained, the dashed
lines in the one-dimensional distributions are the 5% and 95% quantiles,
and contours are at 1σ, 2σ and 3σ.
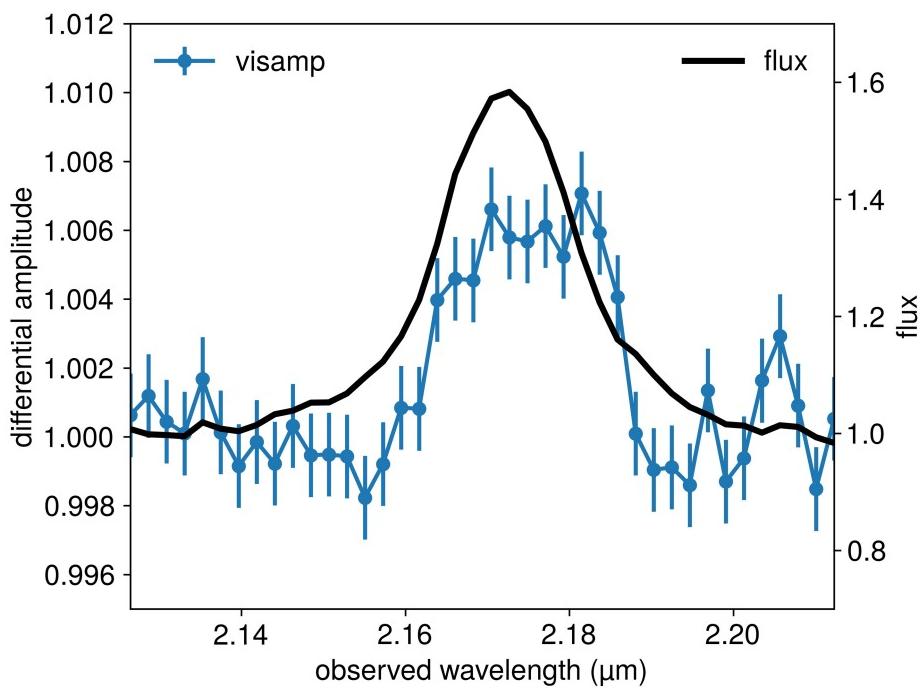
Extended Data Fig. 4 | Additional BLR size constraints. We averaged the differential visibility amplitude (blue, 'visamp') for 3C 273 over all epochs and between the two longest baselines (UT4-1 and UT3-1). Error bars are 1σ. The amplitude increases at the spectral line (black), demonstrating that the hot dust continuum (which has a radius of about 150 μas) is much larger in size than the BLR. This result rules out a previous claim of a large region size in 3C 273[12], and is consistent with the compact size ($R_{BLR} \approx 50$ μas) that we find independently from modelling interferometric phase and spectral-line data.

Extended Data Table 1 | GRAVITY data used for this work

Date	Mode Resolution/Polarisation	Exposure time on 3C 273 (min)	Seeing (")	Coherence time Tau$_0$ (ms)
2017 Jul 07	MED/COMBINED	40	0.44 – 0.77	4.6 – 6.5
2017 Jul 08	MED/COMBINED	35	0.45 – 0.58	2.8 – 3.8
2018 Jan 08	MED/COMBINED	40	0.44 – 0.59	6.9 – 9.0
2018 Mar 29	MED/COMBINED	25	0.41 – 0.48	11.6 – 15.0
2018 Apr 01	MED/COMBINED	30	0.59 – 1.0	3.7 – 4.7
2018 Apr 02	MED/COMBINED	30	0.46 – 0.85	3.2 – 4.9
2018 May 29	MED/COMBINED	100	0.52 – 0.75	3.1 – 5.8
2018 May 30	MED/COMBINED	90	0.48 – 0.68	2.9 – 4.1

LETTER

Quantum control of surface acoustic-wave phonons

K. J. Satzinger[1,2], Y. P. Zhong[2], H.-S. Chang[2], G. A. Peairs[1,2], A. Bienfait[2], Ming-Han Chou[2,3], A. Y. Cleland[2], C. R. Conner[2], É. Dumur[2,4], J. Grebel[2], I. Gutierrez[2], B. H. November[2], R. G. Povey[2,3], S. J. Whiteley[2,3], D. D. Awschalom[2,4], D. I. Schuster[3] & A. N. Cleland[2,4]*

One of the hallmarks of quantum physics is the generation of non-classical quantum states and superpositions, which has been demonstrated in several quantum systems, including ions, solid-state qubits and photons. However, only indirect demonstrations of non-classical states have been achieved in mechanical systems, despite the scientific appeal and technical utility of such a capability[1,2], including in quantum sensing, computation and communication applications. This is due in part to the highly linear response of most mechanical systems, which makes quantum operations difficult, as well as their characteristically low frequencies, which hinder access to the quantum ground state[3–7]. Here we demonstrate full quantum control of the mechanical state of a macroscale mechanical resonator. We strongly couple a surface acoustic-wave[8] resonator to a superconducting qubit, using the qubit to control and measure quantum states in the mechanical resonator. We generate a non-classical superposition of the zero- and one-phonon Fock states and map this and other states using Wigner tomography[9–14]. Such precise, programmable quantum control is essential to a range of applications of surface acoustic waves in the quantum limit, including the coupling of disparate quantum systems[15,16].

Linear resonant systems are traditionally challenging to control at the level of single quanta because they are always in the correspondence limit[17], where quantum behaviour is indistinguishable from classical motion. The recent advent of engineered quantum devices in the form of qubits has enabled full quantum control over some linear systems, in particular electromagnetic resonators[13,14]. A number of experiments have demonstrated that qubits may provide similar control over mechanical degrees of freedom, including qubits coupled to bulk acoustic modes[3,7,18], surface acoustic waves (SAWs)[19–21] and flexural modes in suspended beams[22–25]. In addition, several experiments have studied entanglement between remote mechanical modes generated via heralding measurements[18,26] and reservoir engineering[27]. Of particular note are experiments in which a superconducting qubit is coupled via a piezoelectric material to a microwave-frequency bulk acoustic mode[28], where the ground state can be achieved at moderate cryogenic temperatures; such experiments include controlled vacuum Rabi swaps between the qubit and the mechanical mode[3,7]. However, the level of quantum control and measurement has been limited by the difficulty in engineering a single mechanical mode with sufficient coupling and quantum state lifetime. More advanced operations, such as synthesizing arbitrary acoustic quantum states and measuring those states using Wigner tomography, remain a challenge. Here we report an important advance in the level of quantum control of a mechanical device, where we couple a superconducting qubit to a microwave-frequency SAW resonance, demonstrating ground-state operation, vacuum Rabi swaps between the qubit and the acoustic mode, and the synthesis of mechanical Fock states as well as a Fock state superposition. We map out the Wigner function for these mechanical states using qubit-based Wigner tomography. We note that a similar achievement has been recently reported in an experiment coupling a superconducting qubit to a bulk acoustic mode[29].

The device that we use for this experiment is shown in Fig. 1. The superconducting qubit is a frequency-tunable planar transmon[30,31], connected to the SAW device through a tunable inductor network that provides electronic control[32] of the coupling strength g_0 (see Supplementary Information). Qubit rotations about the X and Y axes in the Bloch sphere representation are performed using pulses on the microwave (XY) line, and Z-axis rotations are achieved by application of a flux bias current on the frequency-control (Z) line. We measure the qubit state using a dispersively coupled readout resonator (see Supplementary Information). The superconducting qubit is fabricated on a sapphire substrate with standard techniques (see Supplementary Information). The SAW resonator is fabricated separately on a lithium niobate substrate, a strong piezoelectric material commonly used for SAW devices[8]. The SAW resonator comprises an interdigital transducer placed between two Bragg mirrors, designed to support a single SAW resonance in the mirror stop band[8] (see Supplementary Information). The SAW wavelength λ is set by the period of the metal lines that constitute the resonator; here, $\lambda = 1\,\mu m$, which corresponds to a frequency of 4.0 GHz. At the experiment temperature, about 10 mK, both the SAWs and the qubit should be in their quantum ground states. The electromechanical properties of the SAW resonator are modelled using an equivalent electrical circuit with a complex, frequency-dependent acoustic admittance[8] $Y_a(\omega)$ connected in parallel with an interdigital capacitance $C_t = 0.75$ pF. The admittance includes the complete response of the SAW transducer and the interaction of the SAW with the mirrors. The strong electromechanical coupling coefficient of lithium niobate makes it feasible to strongly couple the SAW resonance to a standard transmon-style qubit (see Supplementary Information). The separate qubit and SAW-resonator chips are connected together in a flip-chip assembly, in which the lithium niobate chip is inverted, aligned and affixed to the sapphire chip, and are separated vertically by about 7 μm (see Supplementary Information). Coupling between the two chips is achieved using two overlaid planar inductors, one on each chip. The coupling strength is controlled using a radio-frequency superconducting quantum interference device (SQUID) tunable coupler[32], where an externally controlled flux bias Φ controls the path of the qubit current. We note that the flip-chip technique used here enables a wide range of future hybrid combinations of different substrate types with superconducting or other types of qubits; as shown below, the coherence of the qubit in this experiment was not affected by this approach.

We use qubit measurements to evaluate the SAW resonator. The qubit itself has a lifetime of $T_1 \approx 20\,\mu s$ and a Ramsey lifetime of $T_{2,\mathrm{Ramsey}} \approx 2\,\mu s$ over the frequency range 3.5–4.5 GHz, measured with the coupling g_0 set to zero (see Supplementary Information). Adjusting g_0 away from zero shortens the qubit lifetime and makes it strongly frequency-dependent, as the transducer converts electromagnetic energy from the qubit into acoustic waves. In Fig. 2, we demonstrate this with $|g_0|/2\pi$ set to 2.3 ± 0.1 MHz (all uncertainties are one standard deviation), where acoustic loss is the dominant decay channel for the qubit. We measure the qubit lifetime T_1 as a function of qubit frequency, $\omega_{ge}/2\pi$, and use it to obtain the quality factor $Q = \omega_{ge}T_1$ and

[1]Department of Physics, University of California, Santa Barbara, CA, USA. [2]Institute for Molecular Engineering, University of Chicago, Chicago, IL, USA. [3]Department of Physics, University of Chicago, Chicago, IL, USA. [4]Institute for Molecular Engineering and Materials Science Division, Argonne National Laboratory, Argonne, Lemont, IL, USA. *e-mail: anc@uchicago.edu
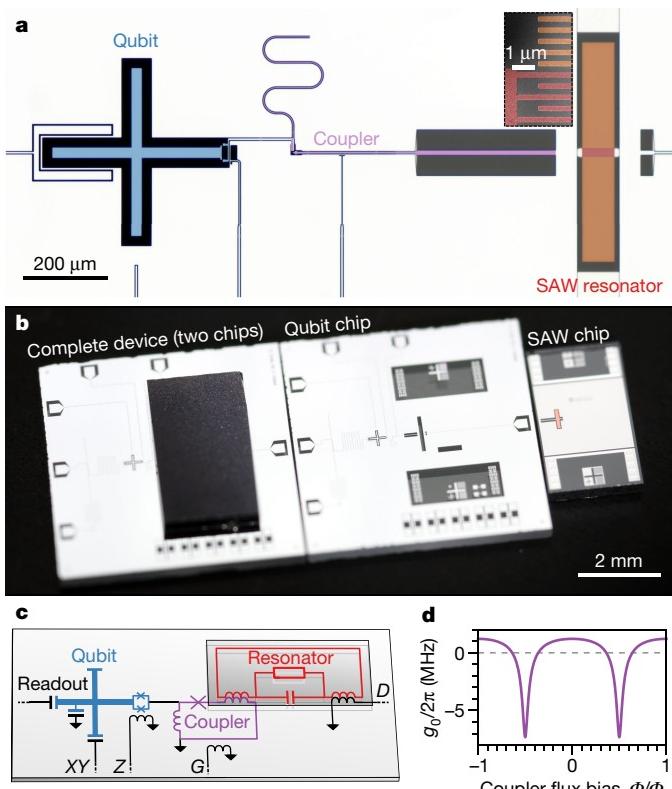
Fig. 1 | Device description. a, False-colour optical photograph of the qubit (left; blue) and SAW resonator (right; red, transducer; orange, mirrors) connected via a tunable coupler (centre; purple). The qubit and coupler are built on a sapphire substrate, whereas the SAW resonator is on a separate lithium niobate substrate. In the figure, the device is viewed from below (through the transparent sapphire substrate). The inset shows a false-colour scanning electron micrograph of the SAW resonator; red, upper left corner of the transducer; orange, mirror. **b**, Photograph showing the flip-chip assembly. Left, the complete device. Centre, the 6 mm × 6 mm sapphire chip. The qubit and coupler are visible near the centre of chip, with control wiring extending to the perimeter. Right, the 2 mm × 4 mm lithium niobate chip with the SAW resonator (red), connected to coupling inductors (horizontal lines). **c**, Circuit diagram. The microwave *XY* line excites the qubit, the *Z* line controls the qubit frequency, the *G* line controls the coupler, and the *D* line coherently displaces the resonator state. The qubit, coupler and control lines are on one plane. The SAW resonator is on a separate chip, represented by the small grey rectangle. Overlaid inductors are mutually coupled. **d**, Qubit–resonator coupling, $g_0/2\pi$, calculated for a range of coupler flux bias values Φ, where $\Phi_0 = h/2e$ is the magnetic flux quantum (h, Planck constant), using the linear circuit model shown in **c** with the experimental parameters (see Supplementary Information).

the corresponding loss $1/Q$. We compare our measurements to the results of a numerical model[8] based on the SAW resonator design with parameters fine-tuned to reproduce the frequency response observed in the qubit loss (see Supplementary Information). The SAW transducer itself can efficiently emit phonons over a wide range of frequencies, roughly from 3.8 GHz to 4.1 GHz, owing to its small number of finger pairs[8] (20 pairs). The SAW mirror reflects acoustic waves efficiently in the mirror stop band from 3.96 GHz to 4.04 GHz. The resultant interference frustrates the transducer emission except when a resonance condition is met, in this case at the single SAW resonance frequency of $\omega_r/2\pi = 3.985$ GHz. The resonator admittance near that resonance can be approximated by an equivalent resonant electrical circuit, which constitutes the Butterworth–van Dyke model[8]. Outside the mirror stop band, the mirror reflection decreases rapidly, and the transducer is free to emit travelling phonons. The qubit sees this as increased loss, especially from 3.85 GHz to 3.90 GHz, where the transducer is most efficient. The ripples in the out-of-band mirror reflection arise from the finite extent of each mirror (500 lines). These features are clearly
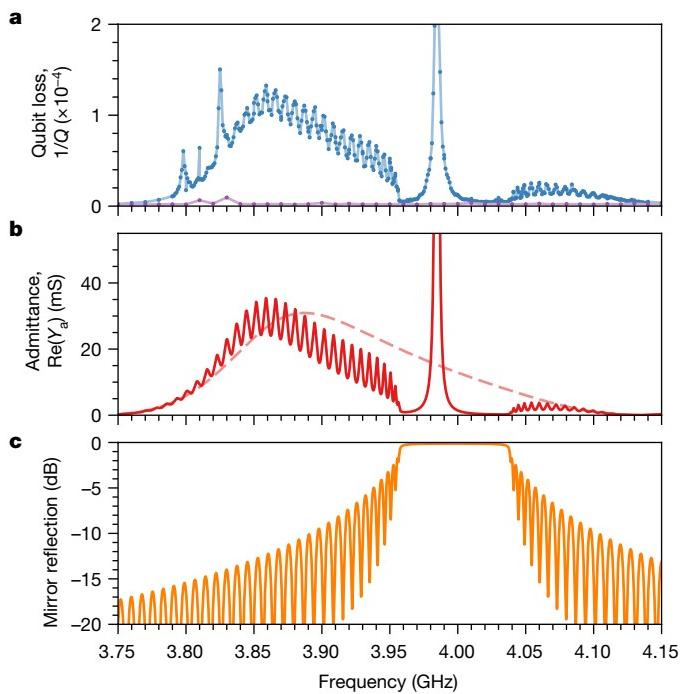
Fig. 2 | Characterization and modelling of SAW admittance. a, Measured qubit loss $1/Q$ as a function of qubit frequency $\omega_{ge}/2\pi$. Blue, $|g_0|/2\pi = 2.3 \pm 0.1$ MHz. The purple line corresponds to minimized g_0. Each data point represents 50,000–100,000 measurements. **b**, Real part of the acoustic admittance of the SAW resonator, $Re(Y_a)$, calculated with a numerical model (see Supplementary Information). The red solid line shows the admittance of the full resonator model; the SAW resonance is the large peak at 3.985 GHz. The pink dashed line is the admittance calculated for the transducer alone, without the mirror structure. **c**, Magnitude of the acoustic reflection coefficient of the mirror model.

displayed in the measured qubit loss. The qubit also weakly couples to unidentified resonances near 3.8 GHz. The SAW resonance at 3.985 GHz can resonantly and rapidly exchange energy with the qubit. In subsequent experiments, we avoid unwanted qubit loss by usually keeping the coupling small and only increasing it when deliberately interacting with the SAW resonance.

We now focus on the interaction between the single SAW resonance and the qubit. In Fig. 3a, we illustrate the full range of qubit coupling to the resonance, determined using spectroscopic measurements of the qubit. We observe a maximum coupling of $|g_0|/(2\pi) = 7.3 \pm 0.1$ MHz, which is equal to half of the avoided-crossing splitting. The ratio of the maximum to the minimum coupling strength is measured to be at least 300 (see Supplementary Information). Figure 3c shows time-domain Rabi swapping of a single excitation between the qubit and the mechanical mode, which represents a photon–phonon exchange in each half-oscillation. A resonant swap operation is executed by setting the qubit frequency to ω_r and turning on the coupling for approximately 37 ns. The number and amplitude of the swaps is primarily limited by the resonator lifetime, T_{1r}.

We show the characterization results for the single-phonon properties of the resonator in Fig. 3c. We prepare a quantum state in the qubit, swap it into the resonator, wait for a delay time t, swap the state back into the qubit, and measure the qubit. The decay of the phonon is consistent with an energy lifetime of $T_{1r} = 148 \pm 1$ ns and a dephasing time of $T_{2r} = 293 \pm 1$ ns, where the ratio $T_{2r}/T_{1r} \approx 2$ is consistent with little to no additional phase decoherence, as expected for a harmonic oscillator. The T_{2r} experiment involves generating a quantum superposition of the resonator phonon Fock states $|0\rangle$ and $|1\rangle$ by performing a Rabi swap from a qubit in the state $(|g\rangle - i|e\rangle)/\sqrt{2}$, where $|g\rangle$ and $|e\rangle$ are the qubit ground and excited states, respectively. The probabilities oscillate at the idle detuning frequency $\Delta/2\pi = 53$ MHz, exhibiting interference between the resonator state and the qubit tomography pulses.

a

b

c
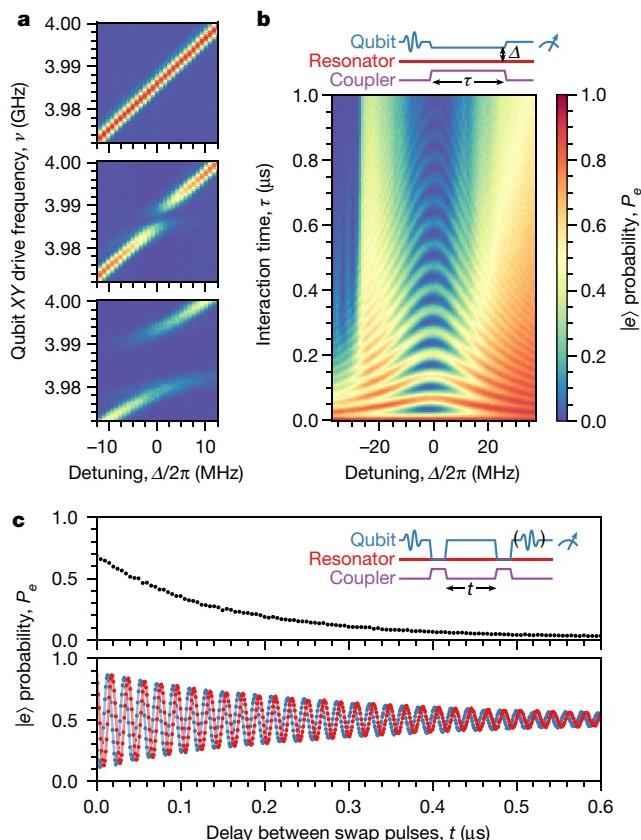
Fig. 3 | Qubit interaction with a single mechanical mode. a, Qubit spectroscopy under three different coupler settings: top, minimum coupling; middle, moderate coupling, $|g_0|/(2\pi) = 2.3 \pm 0.1$ MHz; bottom, maximum coupling, $|g_0|/2\pi = 7.3 \pm 0.1$ MHz. The probability P_e of the excited qubit state $|e\rangle$ is plotted (colour scale as in **b**), measured with the qubit biased at frequency $\omega_r + \Delta$ and driven with a 500-ns-long pulse at frequency ν. Each pixel represents 1,000 repeated measurements. **b**, Results of Rabi-swap experiment. The qubit is excited to $|e\rangle$, and then it is biased to frequency $\omega_r + \Delta$ while the coupling strength is maximized. The qubit and resonator interact for a time τ, and the qubit state is then measured. We plot the probability P_e versus the detuning Δ and the interaction time τ. Each pixel represents 3,000 repeated measurements. **c**, Results of single-phonon experiments using the pulse sequence shown in the inset. Top, measurement of the resonator lifetime, T_{1r}. The qubit is excited to $|e\rangle$ and that excitation is swapped into the resonator. Following a delay time t, the state is swapped back into the qubit, and the qubit is measured. Each data point represents 10,000 repetitions. Bottom, measurement of the dephasing time, T_{2r}. The qubit is excited to $(|g\rangle - i|e\rangle)/\sqrt{2}$, which is swapped into the resonator; after a delay time t the state is swapped back into the qubit. We then conduct qubit tomography, using a second qubit pulse (blue, $X_{\pi/2}$; red, $Y_{\pi/2}$; see Supplementary Information) followed by the qubit measurement. Each data point represents 20,000 repetitions.

We attempt to create the higher Fock state $|2\rangle$ in the SAW resonator by exciting the qubit and swapping its excitation into the resonator two times. We show the result in Fig. 4. The experiment is limited by the resonator lifetime T_{1r}, which is comparable to the duration of the pulse sequence used to generate $|2\rangle$, about 100 ns. We do observe higher-frequency oscillations in the initial interaction, as expected. The experimental result is in excellent agreement with a numerical master-equation model, which is fitted to the experiment by adjusting the initial qubit and resonator states. The resonator state is closest to $|2\rangle$ after an interaction time of $\tau = 26$ ns. At that time, the resonator state calculated by the model is a statistical mixture of 47.3% $|2\rangle$, 38.2% $|1\rangle$ and 14.5% $|0\rangle$, with the unwanted lower states appearing owing to decay during state preparation.

We now characterize the quantum state of the resonator in greater detail. Verifying that the resonator is indeed in its ground state is an

a **b**
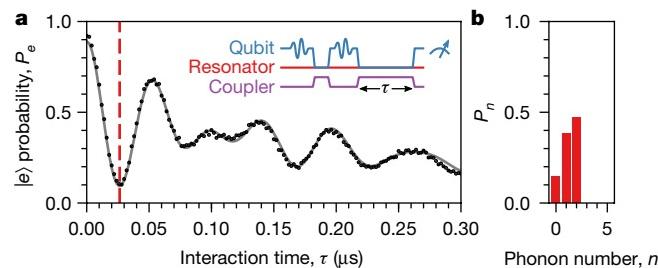
Fig. 4 | Generation of the $|2\rangle$ state. a, Qubit evolution, nominally starting with the qubit in $|e\rangle$ and the resonator in $|1\rangle$. Black points, experimental results; grey line, numerical model; red dashed line, time when the resonator state is closest to $|2\rangle$. The inset shows the pulse sequence used in the experiment: the qubit is excited to $|e\rangle$, which is swapped into the resonator; the qubit is again excited to $|e\rangle$, and then it interacts with the resonator for time τ. Each point represents 3,000 repetitions. **b**, The phonon number probability distribution, P_n, calculated by the model at the time indicated by the red dashed line in **a** ($\tau = 26$ ns).

important step in evaluating its quantum behaviour. We examine the residual thermal populations in the qubit and resonator excited states, $|e\rangle$ and $|1\rangle$, respectively, using a Rabi population measurement technique[7,33] (see Supplementary Information). Driven transitions between $|e\rangle$ and the second excited qubit state, $|f\rangle$, are used to quantify the $|e\rangle$ population by measuring the amplitudes of Rabi-like oscillations. The experimental results are shown in Fig. 5a, where we vary the amplitude of a microwave pulse that drives e–f transitions. In the left panel, we show the result of probing the ground-state population of the qubit; the large-amplitude oscillations show near-unity initial ground-state population. In the right panel, we probe the excited-state population, which is much smaller. We calculate the excited-state population from the amplitudes of these oscillations (see Supplementary Information). When performing the experiment on the qubit alone, we observe an excited-state population of 0.0169 ± 0.0002. To assess the thermal population of the resonator, we first execute a resonant-swap operation, and then we conduct the experiment again. The swap exchanges the small excited-state populations in the resonator and the qubit. In this case, we observe an excited-state population of 0.0049 ± 0.0002, which we interpret as an upper bound on the excited-state population of the resonator[7].

The level of control achievable in this experiment allows us to controllably generate the resonator states $|0\rangle$, $|1\rangle$, $(|0\rangle + |1\rangle)/\sqrt{2}$ and, to a lesser extent, $|2\rangle$. We prepare these resonator states deterministically, by exciting the qubit and transferring energy into the resonator with resonant swaps. We use Wigner tomography to determine the fidelities of these quantum states[13] (see Supplementary Information), examining the three lowest-energy states in detail. Following state preparation, we measure the Wigner function $W(\alpha)$ of the resonator by using the qubit to measure the parity of the resonator states at different complex displacements α in the resonator phase space (see Supplementary Information). The required displacements α are created by driving the resonator with a resonant Gaussian microwave pulse applied to a control line (see Fig. 1c). During the pulse, the coupling is turned off, and the qubit is detuned above the resonator by $\Delta/2\pi = 400$ MHz.

With the qubit initially in its ground state $|g\rangle$, we allow the qubit and resonator to resonantly interact for a time τ, and then we measure the qubit. An example is shown in Fig. 5b. The plot of the qubit state as a function of delay τ contains information about the displaced-resonator state. We fit the experimental results with a numerical master-equation model to deduce the phonon number (n) distribution of the displaced resonator, P_n. We then calculate the Wigner function, which is proportional to the phonon number parity[13] (see Supplementary Information). We repeat the experiment for many values of α to map out the Wigner function. The results are displayed in Fig. 5d, along with the prediction of the numerical model using the same pulse sequence. The value of each pixel is determined independently. We then convert each experimental $W(\alpha)$ into a density matrix ρ (see Supplementary Information).
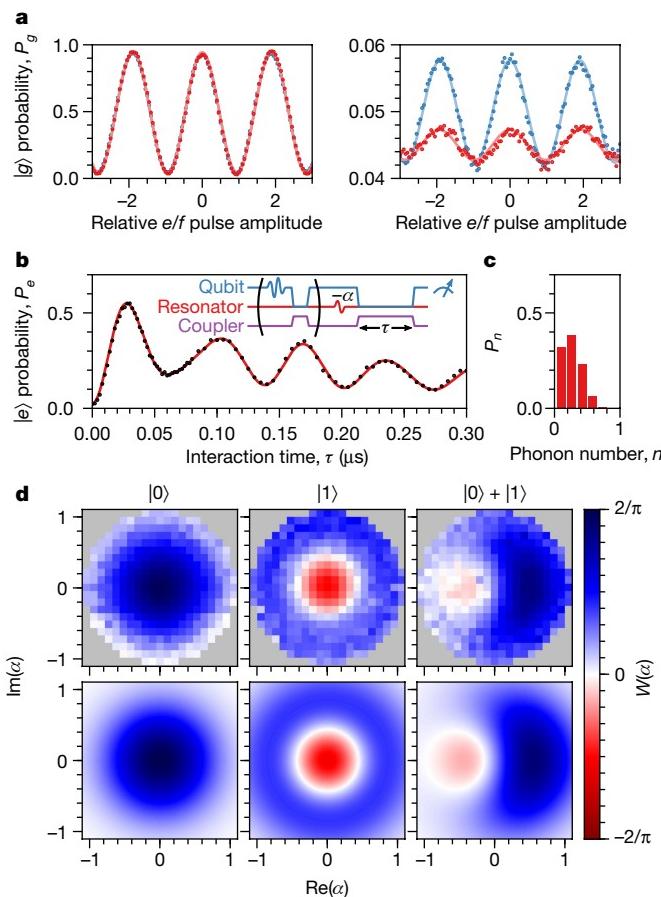
Fig. 5 | Resonator state characterization. a, Results of Rabi population measurement[33], performed to determine the steady-state $|e\rangle$ qubit population by driving transitions between $|e\rangle$ and $|f\rangle$ (see Supplementary Information). The sequences used to probe the ground-state (left; 10,000 repetitions per data point) and excited-state (right; 200,000 repetitions per point) populations are applied to the equilibrium qubit state (blue) and the qubit after swapping with the resonator (red)[7]. The e–f pulse amplitude is normalized to the amplitude that swaps $|e\rangle$ and $|f\rangle$, and a negative pulse amplitude means that the pulse phase is π. **b**, Example Wigner tomography result showing the evolution of the qubit as it interacts with a displaced-resonator $|1\rangle$ state (black points). The red line is a fit (see **c**). The inset shows the synthesis of the mechanical state, which—for states other than the ground state—is indicated by the sequence in parantheses, and the pulse sequence used in Wigner tomography. The resonator state is displaced with coherent amplitude $-\alpha$. The qubit interacts with the displaced resonator state for a time τ before it is measured. Each point represents 3,000 repetitions. **c**, Example fit of the phonon number distribution P_n (statistical uncertainty, 0.004) obtained from the experimental results shown in **b**. **d**, Wigner functions $W(\alpha)$ of the SAW resonator quantum states. Top, experimental results; the $(|0\rangle + |1\rangle)/\sqrt{2}$ Wigner function is rotated by 90° to compensate for relative phase accumulation during the pulse sequence[13]. Each pixel represents 255,000 repetitions. Bottom, result of the numerical model.

From the density matrices, we calculate the quantum state fidelities to the ideal states $|\psi\rangle$, $F = \sqrt{\langle\psi|\,\rho\,|\psi\rangle}$. We obtain $F = 0.985 \pm 0.005$ for $|0\rangle$, $F = 0.858 \pm 0.007$ for $|0\rangle$ and $F = 0.945 \pm 0.006$ for $(|0\rangle + |1\rangle)/\sqrt{2}$. The numerical model predicts similar fidelities: $F = 0.998$, 0.879 and 0.962, respectively. These experiments would benefit from a longer phonon lifetime T_{1r} and larger coupling strength.

In conclusion, we demonstrate high-fidelity, on-demand synthesis of quantum states in a macroscale mechanical resonator and characterize them with Wigner tomography. The primary limitation in these experiments is the phonon lifetime in combination with the maximum coupling strength. These could be improved substantially in future work, for example, with design and material changes in the mechanical

resonator and adjustments to the coupling circuit. Our demonstration involves a hybrid architecture incorporating a high-performance qubit with strong tunable coupling to SAWs. This scalable platform holds promise for future quantum acoustics experiments coupling stationary qubits to 'flying' qubits based on phonons. The technologies demonstrated here may also enable a wide range of experiments coupling superconducting circuits to diverse quantum systems, such as semiconductor spin systems.

Data availability

The datasets supporting this work are available from the corresponding author on request.

1. Stannigel, K., Rabl, P., Sørensen, A. S., Zoller, P. & Lukin, M. D. Optomechanical transducers for long-distance quantum communication. *Phys. Rev. Lett.* **105**, 220501 (2010).
2. Kurizki, G. et al. Quantum technologies with hybrid systems. *Proc. Natl Acad. Sci. USA* **112**, 3866–3873 (2015).
3. O'Connell, A. D. et al. Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703 (2010).
4. Teufel, J. D. et al. Sideband cooling of micromechanical motion to the quantum ground state. *Nature* **475**, 359–363 (2011).
5. Chan, J. et al. Laser cooling of a nanomechanical oscillator into its quantum ground state. *Nature* **478**, 89–92 (2011).
6. Wollman, E. E. et al. Quantum squeezing of motion in a mechanical resonator. *Science* **349**, 952–955 (2015).
7. Chu, Y. et al. Quantum acoustics with superconducting qubits. *Science* **358**, 199–202 (2017).
8. Morgan, D. *Surface Acoustic Wave Filters: With Applications to Electronic Communications and Signal Processing* 2nd edn (Academic Press, Oxford, 2007).
9. Law, C. K. & Eberly, J. H. Arbitrary control of a quantum electromagnetic field. *Phys. Rev. Lett.* **76**, 1055–1058 (1996).
10. Banaszek, K., Radzewicz, C., Wódkiewicz, K. & Krasiński, J. S. Direct measurement of the Wigner function by photon counting. *Phys. Rev. A* **60**, 674–677 (1999).
11. Bertet, P. et al. Direct measurement of the Wigner function of a one-photon Fock state in a cavity. *Phys. Rev. Lett.* **89**, 200402 (2002).
12. Haroche, S. & Raimond, J.-M. *Exploring the Quantum: Atoms, Cavities and Photons* (Oxford Univ. Press, Oxford, 2006).
13. Hofheinz, M. et al. Synthesizing arbitrary quantum states in a superconducting resonator. *Nature* **459**, 546–549 (2009).
14. Vlastakis, B. et al. Deterministically encoding quantum information using 100-photon Schrödinger cat states. *Science* **342**, 607–610 (2013).
15. Schuetz, M. et al. Universal quantum transducers based on surface acoustic waves. *Phys. Rev. X* **5**, 031031 (2015).
16. Whiteley, S. J. et al. Probing spin–phonon interactions in silicon carbide with Gaussian acoustics. Preprint at https://arxiv.org/abs/1804.10996 (2018).
17. Bohr, N. Über die Serienspektra der Elemente. *Z. Phys.* **2**, 423–469 (1920).
18. Lee, K. C. et al. Entangling macroscopic diamonds at room temperature. *Science* **334**, 1253–1256 (2011).
19. Gustafsson, M. V. et al. Propagating phonons coupled to an artificial atom. *Science* **346**, 207–211 (2014).
20. Manenti, R. et al. Circuit quantum acoustodynamics with surface acoustic waves. *Nat. Commun.* **8**, 975 (2017).
21. Moores, B. A., Sletton, L. R., Viennot, J. J. & Lehnert, K. W. Cavity quantum acoustic device in the multimode strong coupling regime. *Phys. Rev. Lett.* **120**, 227701 (2018).
22. Arcizet, O. et al. A single nitrogen-vacancy defect coupled to a nanomechanical oscillator. *Nat. Phys.* **7**, 879–883 (2011).
23. Kolkowitz, S. et al. Coherent sensing of a mechanical resonator with a single-spin qubit. *Science* **335**, 1603–1606 (2012).
24. Yeo, I. et al. Strain-mediated coupling in a quantum dot–mechanical oscillator hybrid system. *Nat. Nanotechnol.* **9**, 106–110 (2013).
25. Lee, K. W. et al. Strain coupling of a mechanical resonator to a single quantum emitter in diamond. *Phys. Rev. Appl.* **6**, 034005 (2016).
26. Riedinger, R. et al. Remote quantum entanglement between two micromechanical oscillators. *Nature* **556**, 473–477 (2018).
27. Ockeloen-Korppi, C. F. et al. Stabilized entanglement of massive mechanical oscillators. *Nature* **556**, 478–482 (2018).
28. Cleland, A. N. & Geller, M. R. Superconducting qubit storage and entanglement with nanomechanical resonators. *Phys. Rev. Lett.* **93**, 070501 (2004).
29. Chu, Y. et al. Creation and control of multi-phonon Fock states in a bulk acoustic-wave resonator. *Nature* https://doi.org/10.1038/s41586-018-0717-7 (2018).
30. Koch, J. et al. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
31. Barends, R. et al. Coherent Josephson qubit suitable for scalable quantum integrated circuits. *Phys. Rev. Lett.* **111**, 080502 (2013).
32. Chen, Y. et al. Qubit architecture with high coherence and fast tunable coupling. *Phys. Rev. Lett.* **113**, 220502 (2014).

33. Geerlings, K. et al. Demonstrating a driven reset protocol for a superconducting qubit. *Phys. Rev. Lett.* **110**, 120501 (2013).

Author contributions K.J.S. designed and fabricated the devices. K.J.S., H.-S.C., J.G., A.Y.C. and S.J.W. developed the fabrication processes. G.A.P., É.D. and A.N.C. contributed to device design. K.J.S. performed the experiments and analysed the data with assistance from Y.P.Z., A.B. and É.D. Assistance was provided by I.G. and B.H.N. A.N.C., D.I.S. and D.D.A. advised on all efforts. All authors contributed to discussions and the production of the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0719-5.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.N.C.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Creation and control of multi-phonon Fock states in a bulk acoustic-wave resonator

Yiwen Chu[1,2]*, Prashanta Kharel[1,2], Taekwan Yoon[1,2], Luigi Frunzio[1,2], Peter T. Rakich[1,2] & Robert J. Schoelkopf[1,2]*

Quantum states of mechanical motion can be important resources for quantum information, metrology and studies of fundamental physics. Recent demonstrations of superconducting qubits coupled to acoustic resonators have opened up the possibility of performing quantum operations on macroscale motional modes[1–3], which can act as long-lived quantum memories or transducers. In addition, they can potentially be used to test decoherence mechanisms in macroscale objects and other modifications to standard quantum theory[4,5]. Many of these applications call for the ability to create and characterize complex quantum states, such as states with a well defined phonon number, also known as phonon Fock states. Such capabilities require fast quantum operations and long coherence times of the mechanical mode. Here we demonstrate the controlled generation of multi-phonon Fock states in a macroscale bulk acoustic-wave resonator. We also perform Wigner tomography and state reconstruction to highlight the quantum nature of the prepared states[6]. These demonstrations are made possible by the long coherence times of our acoustic resonator and our ability to selectively couple a superconducting qubit to individual phonon modes. Our work shows that circuit quantum acoustodynamics[7] enables sophisticated quantum control of macroscale mechanical objects and opens up the possibility of using acoustic modes as quantum resources.

Light and sound are two examples of familiar wave phenomena in the classical world. Until now, the field of quantum optics has extensively demonstrated the particle nature of light in quantum mechanics through the study of single photons and other non-Gaussian electromagnetic states. The concept of particles of sound, or phonons, is used widely in solid state physics. However, the ability to create states of individual phonons has only been demonstrated in a few instances[1,3,8], whereas the complete quantum tomography of such states has only been achieved in a single trapped ion[6]. This disparity between electromagnetic and acoustic degrees of freedom is largely due to sound propagating inside the complex and potentially lossy environment of a massive material, rather than in vacuum. As a result, an open question remains: is it feasible to control and measure complex quantum states in the motion of a macroscale solid-state object, or what we usually think of as sound, analogously to what has been done with light?

The relatively new field of quantum acoustics attempts to answer this question using a variety of optomechanical and electromechanical systems[1,2,7,9–13], and one particularly promising approach is circuit quantum acoustodynamics (QAD)[1,2,7,13]. In analogy to circuit quantum electrodynamics (QED), circuit QAD uses superconducting quantum circuits that operate at microwave frequencies to manipulate and measure mechanical resonators. Circuit QAD takes advantage of the strong interactions between mechanics and electromagnetism enabled by, for example, piezoelectricity. It also incorporates the nonlinearity provided by the Josephson junction, which is a crucial ingredient for creating non-Gaussian states of motion. In turn, the ability to create these states makes mechanical resonators useful as resources in quantum circuits, offering capabilities beyond those of electromagnetic resonators. For example, mechanical transduction is a promising method

for transferring quantum information between microwave circuits and other systems, such as optical light or spin qubits[14,15]. Owing to the difference between the speeds of sound and light, an acoustic resonator is much more compact and well isolated than an electromagnetic one at the same frequency and provides many more independent modes that are individually addressable by a superconducting qubit. Such an architecture is desirable for simulating many-body quantum systems[2,16] and provides a highly hardware-efficient way of storing, protecting and manipulating quantum information using bosonic encodings[17,18]. These examples show that by repurposing the toolbox of circuit QED through the similarities between light and sound, circuit QAD allows us to make use of the important differences between these quantum degrees of freedom. However, in order to access this toolbox, we first need to demonstrate that a circuit QAD system can be engineered to have the necessary mode structure, strong enough interactions and sufficient quantum coherence to create and characterize quantum states of motion.

In this work, we experimentally prepare and perform full quantum tomography on Fock states of phonons and their superpositions inside a high-overtone bulk acoustic-wave resonator (HBAR). This is enabled by a robust new flip-chip device geometry that couples a superconducting transmon qubit to the HBAR. This geometry allows us to optimize the design of the acoustic resonator and the qubit separately to extend phonon coherence while enhancing the selectivity of the coupling to a single mode. The combination of these improvements leads to a device that is deeper in the strong-coupling regime of circuit QAD, which is necessary for the generation and manipulation of more complex quantum states. We note that a similar demonstration using a superconducting qubit and surface acoustic waves has been recently reported[19].

We now describe the motivation behind the design of our circuit QAD system in more detail. Figure 1a shows a schematic of our device, which we call the ℏBAR. The first important difference from our previous device[3] is the flip-chip geometry, where the qubit and acoustic resonator are now on separate sapphire chips[19]. This simplifies the fabrication procedure and increases the yield of successful devices (see Supplementary Information) while allowing qubits and acoustic resonators to be individually tested before assembly. Second, the ℏBAR incorporates a plano-convex acoustic resonator that is fabricated using a simple, robust method and supports stable, transversely confined acoustic modes (see Supplementary Information). Because the measured acoustic lifetime in the previous unstable resonator geometry[3] was consistent with being limited by diffraction loss, this modification to our device could considerably improve phonon coherence. Another important requirement is the ability to selectively couple the qubit to a single acoustic mode. This is partly achieved by the plano-convex resonator design, which allows us to control the frequency spacing between transverse modes. To further increase mode selectivity, the third improvement is the addition of an optimized transduction electrode to the qubit. The electrode was designed to match the strain profile of the fundamental Gaussian transverse mode of the acoustic resonator (see Supplementary Information). We point out that even though the acoustic resonator is not in physical contact with the electrode, the

[1]Department of Applied Physics, Yale University, New Haven, CT, USA. [2]Yale Quantum Institute, Yale University, New Haven, CT, USA. *e-mail: yiwen.chu@yale.edu; robert.schoelkopf@yale.edu

a

b
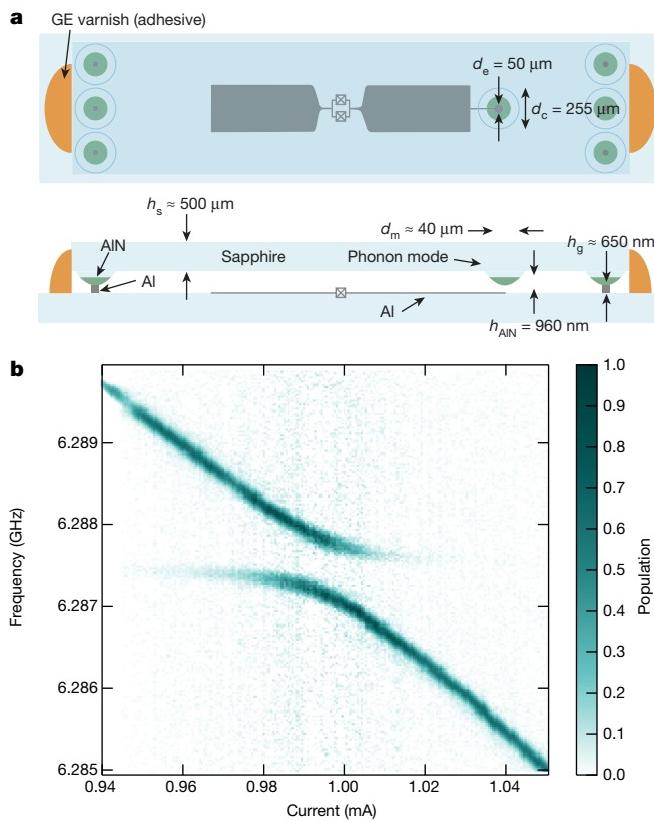
Fig. 1 | The ħBAR device and strong qubit–phonon coupling. a, Top-view (top) and side-view (bottom) schematics of the ħBAR device (not to scale). The chip containing the acoustic resonator has six nominally identical resonators on its edges, with Al (grey) spacers deposited on them. The measured thickness of the Al spacer between the qubit and acoustic resonator (h_g) is shown, but the actual spacing may be larger owing to imperfections in the flip-chip assembly. Other dimensions indicated are the diameters of the transducer electrode (d_e), curved resonator surface (d_c) and acoustic-mode waist (d_m), along with the thicknesses of the AlN (h_{AlN}; green) and sapphire substrate (h_s; blue). **b**, Spectroscopy results of the transmon qubit near the l_1 acoustic mode, measured while varying the current in an external coil used to flux-tune the qubit frequency.

electric field of the qubit extends across the gap between the two chips and through the AlN film, thus allowing piezoelectric transduction.

We now experimentally show that the new design does indeed lead to improvements in the electro-mechanical coupling, acoustic-mode spectrum and coherence of our device. As in our previous work, the ħBAR is measured using a standard circuit QED setup that allows flux-tuning of the qubit frequency using an externally applied magnetic field. Figure 1b shows qubit spectroscopy results near the $l = l_1$ and $m, n = 0$ mode of the ħBAR, which reveal a single distinct anticrossing feature. Here l is the longitudinal mode number and m, n are the mode numbers of the Hermite–Gaussian-like transverse modes. $l_1 \approx 466$ corresponds to the highest-frequency longitudinal mode that is fully within the tunable range of the qubit, as indicated in Fig. 2, where we investigate the mode structure of the ħBAR over several longitudinal free spectral ranges. Figure 2a shows the time dynamics of the qubit–phonon interaction for different qubit frequencies, which reveals features indicative of vacuum Rabi oscillations that are spaced by the longitudinal free spectral range $\nu_{FSR} = 13.5$ MHz. Each oscillatory feature corresponds to an anticrossing similar to the one shown in Fig. 1b. The Fourier transform of the data in Fig. 2a is shown in Fig. 2b and gives a qubit–phonon coupling rate of $g_0 = 2\pi \times (350 \pm 3)$ kHz. In addition to the dominant set of oscillations corresponding to the $m, n = 0$ Gaussian modes, there are clear signatures of other acoustic modes in Fig. 2a, b, which correspond to higher-order transverse modes, as indicated by simulations (see Supplementary Information). However, the closest

observable higher-order mode is about 1 MHz away from the $m, n = 0$ mode and about ten times less strongly coupled to the qubit, while all other higher-order modes are at least five times less strongly coupled. (From now on, we use only the longitudinal mode number to represent the $m, n = 0$ modes.) These results indicate that the ħBAR is a good approximation of a system in which the qubit can be tuned to interact with a single acoustic mode at a time.

We demonstrate the improvements in the coherence of our system by performing quantum operations on the phonon mode using the qubit. Using techniques described in our previous work[3], we find that the phonon mode has a lifetime of $T_1 = 64 \pm 2$ μs, a Ramsey decoherence time of $T_2 = 38 \pm 2$ μs and an echo decoherence time of $T_{2e} = 45 \pm 2$ μs. On other devices, we measured the phonon lifetime to be as long as $T_1 = 113 \pm 4$ μs. These coherence times are comparable to that of state-of-the art superconducting qubits and suggest that the plano-convex resonator design does indeed support much-longer-lived phonons. The qubit in this device has a $T_1 = 7 \pm 1$ μs, which is similar to that of our previous device. As will be discussed later, we believe that these device parameters can be further improved through modifications of the materials, fabrication procedure and device geometry.

The improvements presented above allow us to perform quantum operations on the phonon mode with a new level of sophistication, which we now illustrate by creating and measuring multi-phonon Fock states. We use a procedure for Fock state preparation that has previously only been demonstrated in electromagnetic systems[20] (Fig. 3a). The experiment begins with the qubit set to a frequency ν_0 that is detuned by $\delta = -5$ MHz from the target l_1 phonon mode at frequency ν_1. The qubit ideally starts out in the ground state $|g\rangle$, but in reality has a thermal population of 4%–8% in the excited state $|e\rangle$. The phonon modes, on the other hand, have been shown to be colder[3]. Therefore we first perform a swap operation between the qubit and the l_2 mode with frequency ν_2. This procedure effectively uses an additional acoustic mode to cool the qubit to an excited-state population of about 2%. The qubit is then excited with a π pulse and brought into resonance with the l_1 mode to transfer its energy into the acoustic resonator using a swap operation. This is repeated N times to climb up the Fock state ladder and ideally results in a state of N phonons, which is then probed by bringing the qubit and phonon to resonance for a variable time t and measuring the final qubit state. We note that this measurement procedure gives the total population in the qubit excited-state subspace of the joint system and traces over the resonator state. The resulting time dynamics of the population $p_{e,N}(t)$ for up to $N = 7$ is shown in Fig. 3b. In Fig. 3c, we plot the Fourier transform of the data in Fig. 3b. As expected, we observe oscillations with a dominant frequency of $2g_N = 2\sqrt{N} g_0$, which corresponds to the rate of energy exchange between the $|g, N\rangle$ and $|e, N - 1\rangle$ states.

To characterize the states that we have created more quantitatively, we extract the population in each phonon Fock state n after performing an N-phonon preparation. We do this by first simulating the expected time traces, $p_{e,n}(t)$, assuming that the phonon mode is prepared in an ideal Fock state ranging from $n = 1$ to $n_{max} = 14$. The independently measured value of g_0, along with the qubit and phonon decay and dephasing rates, are used in the simulations. Then, the experimental data for each N (Fig. 3b) are fitted to a weighted sum of the form

$$p_{e,N}(t) = \sum_{n=1}^{n_{max}} p_{n,N} p_{e,n}(t) \tag{1}$$

where $p_{n,N}$ is then the population at $|g, n\rangle$ after performing an N-phonon preparation. The fit for each N is subject to the constraints $p_{n,N} \leq 1 \; \forall \; n$ and $\sum_{n=1}^{n_{max}} p_{n,N} \leq 1$. Finally, the population in the zero-phonon state is calculated as $p_{0,N} = 1 - \sum_{n=1}^{n_{max}} p_{n,N}$. Ideally, $p_{n,N} = \delta_{n,N}$. As shown in Fig. 3d, we observe that the resulting distribution of populations for each experiment is indeed peaked at $n = N$. However, the population in the nominally prepared state decreases with increasing N. We find that $p_{1,1} = 0.86$, which is consistent with a simple estimate that takes into account the energy decay from the one-excitation
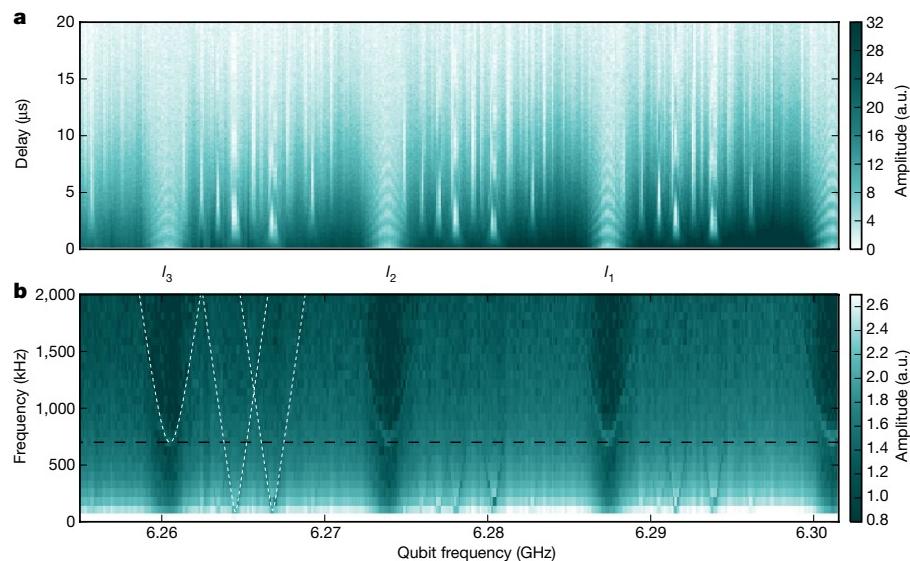
Fig. 2 | Mode structure of the ħBAR. a, Time dynamics of the qubit–phonon interaction, determined by exciting the qubit and measuring its excited-state population after a variable delay as the qubit is flux-tuned (a.u., arbitrary units). **b**, Logarithm of the Fourier transform of the data shown in **a**. The qubit frequencies shown on the horizontal axis are determined from spectroscopy data taken at each applied flux. The three highest-frequency fundamental transverse modes ($m, n = 0$)

that are fully accessible by the qubit are shown and labelled with their longitudinal mode numbers. The white dashed lines indicate the hyperbolic dependence of the effective vacuum Rabi frequencies for the three most dominant modes in one free spectral range. The black dashed line indicates the value of $2g_0$ for the fundamental mode, which is at least a factor of five larger than the coupling rates to the other modes.

manifold during a swap operation, which is dominated by the qubit decay rate, and the imperfect preparation of the qubit in $|g\rangle$. For larger N values, the state preparation may be affected by additional effects, such as off-resonant driving of the phonon mode during the qubit π pulses, which could lead to excess population in the $n > N$ states. We

also find that the largest source of potential error in extracting $p_{n,N}$ comes from uncertainty in the system parameters that are used in simulating $p_{e,n}(t)$. In particular, slight drifts of the qubit frequency can result in a mismatch between the value of $2g_0$ used in the simulations and the actual oscillation frequency of the vacuum Rabi data. An
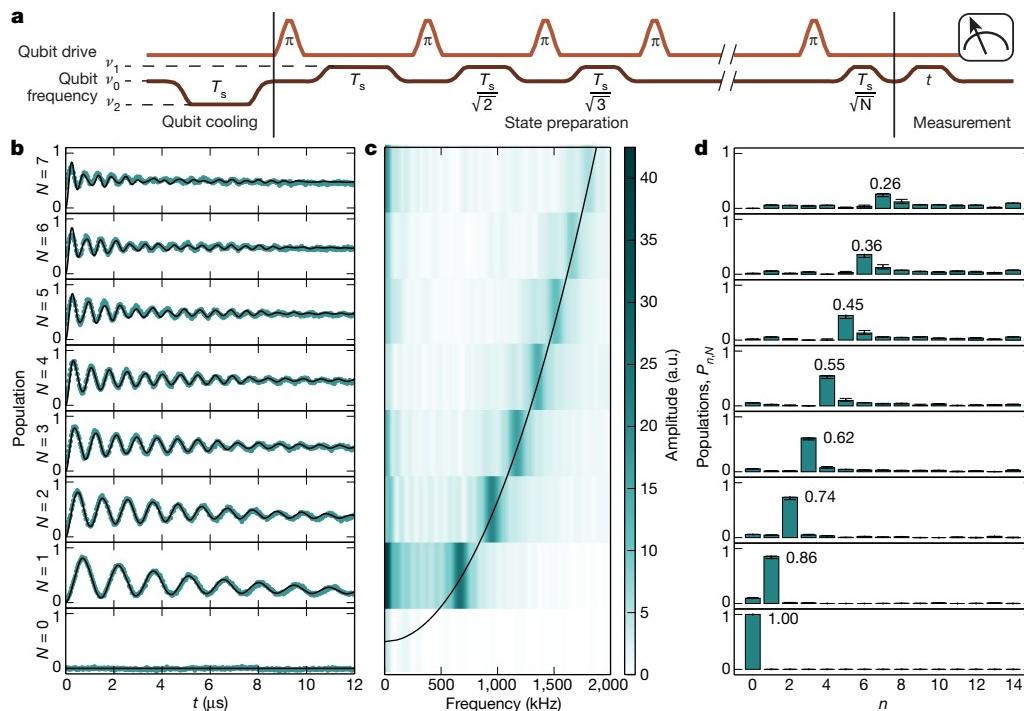
Fig. 3 | Climbing the phonon Fock state ladder. a, Pulse sequence for the generation and measurement of phonon Fock states. $T_s = \pi/2g_0$ is the duration of a swap operation in the one-excitation manifold. In the state preparation step, the duration of the kth swap is scaled to account for the coupling rate $g_k = \sqrt{k}g_0$ between the $|g, k\rangle$ and $|e, k-1\rangle$ states. Pulses intended to excite the qubit from $|g\rangle$ to $|e\rangle$ are labelled with 'π'. The qubit frequencies ν_0, ν_1 and ν_2 are described in the text. **b**, Qubit excited-state population after interacting with the phonon for a time t following an

N-phonon preparation procedure. Black lines are fits used to extract the Fock state populations shown in **d**. **c**, Fourier transform of the data shown in **b**, obtained by subtracting the mean of each dataset in **b** and appending copies of the resultant final value to effectively smooth the Fourier transform. The black line corresponds to $2g_N$. **d**, Populations in the n Fock state, extracted from **b**. The numbers show the populations in $n = N$. Error bars indicate the result of changing the value of g_0 in the simulations of $p_{e,n}(t)$ by ± 5 kHz (see main text).
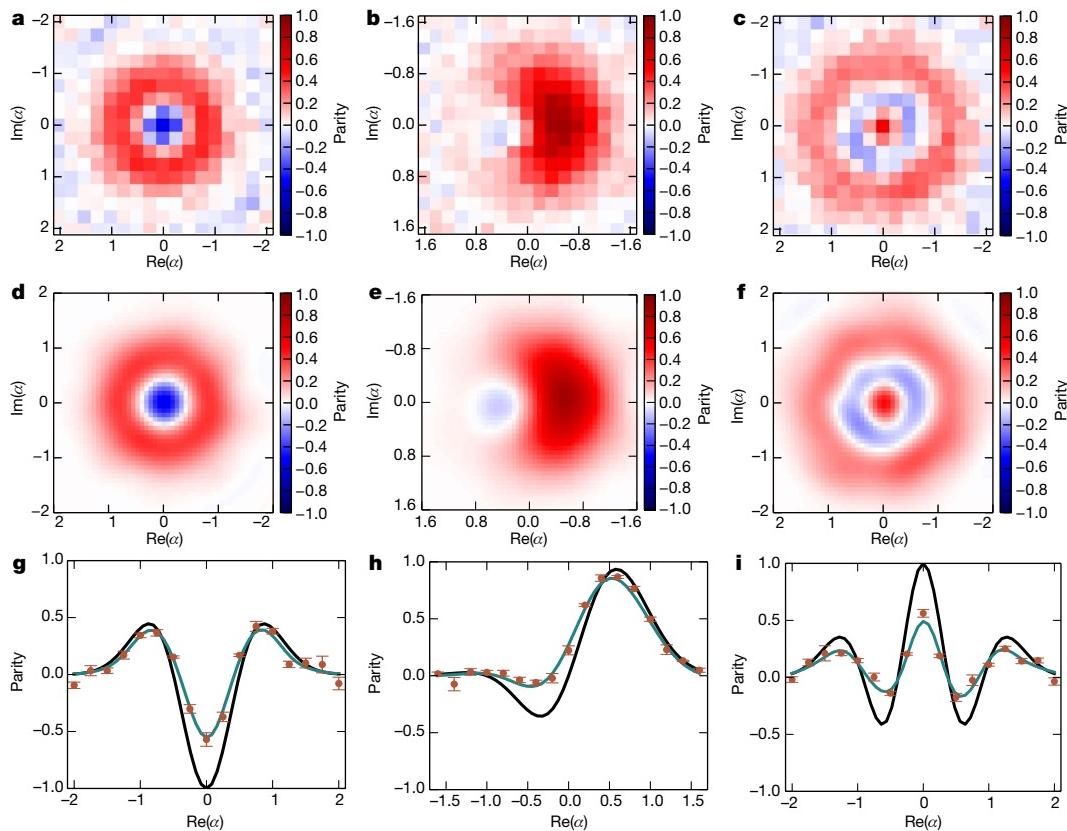
Fig. 4 | Wigner tomography of non-classical states of motion.
a–c, Measured Wigner functions of the prepared states $|1\rangle$,
$(|0\rangle + |1\rangle)/\sqrt{2}$ and $|2\rangle$. Each grid point is a separate experiment with
displacement by a phase-space amplitude α. **d–f,** Wigner functions of the
density matrices reconstructed from **a–c**. **g–i,** Cuts of the ideal Wigner
function (black line), data (brown points) and reconstructed Wigner
function (green line) along the $\text{Im}(\alpha) = 0$ axis. Negative values of the
Wigner function are indicators of a non-classical state of motion. The
error bars are extracted in the same way as in Fig. 3d (see main text).

estimate of the effect of such miscalibrations is given by the error bars
in Fig. 3d.

We now build on our ability of extracting the phonon number distri-
bution to perform full Wigner tomography and explore the quantum
nature of the prepared mechanical state. As in previous circuit QED
and trapped-ion experiments[6,21], we use the definition[22]

$$P(\alpha) = \text{tr}[\hat{D}(-\alpha)\rho\hat{D}(\alpha)\hat{P}] = \frac{\pi}{2}W(\alpha) \tag{2}$$

Here $P(\alpha)$ and $W(\alpha)$ are the values of the displaced parity and Wigner
functions, respectively, at a phase-space amplitude α, ρ is the prepared
state and $\hat{P}$ is the parity operator. From now on, we plot the values of
$P(\alpha)$ for clarity, but use the terms 'displaced parity' and 'Wigner func-
tion' interchangeably. The resonator displacement $\hat{D}(\alpha)$ is implemented
by a microwave pulse at the phonon frequency while the qubit is
detuned at ν_0. Under these conditions, the phonon mode is still coupled
to the microwave drive port, in part owing to its hybridization with the
qubit. To verify this and calibrate our displacement amplitudes, we first
apply a Gaussian phonon drive pulse of varying amplitude α with a
root-mean-square width of 1 μs and truncated to a total length of 4 μs.
We then measure the subsequent Fock state populations $p_{n,|0\rangle}(\alpha)$ and
check that they agree well with the expected Poisson distributions up
to an overall scaling between the amplitudes of the applied drive and
the actual displacement (see Supplementary Information). We can then
calculate the displaced parity for the vacuum state $|0\rangle$ using
$P_{|0\rangle}(\alpha) = \sum_n(-1)^n p_{n,|0\rangle}(\alpha)$. Similarly, we can measure the displaced
parity $P_\rho(\alpha)$ for an arbitrary state ρ by adding a phonon drive pulse
between state preparation and measurement.

In Fig. 4, we present the results of Wigner state tomography on the
nominally prepared states $|1\rangle$, $(|0\rangle + |1\rangle)/\sqrt{2}$ and $|2\rangle$. The

$(|0\rangle + |1\rangle)/\sqrt{2}$ state was prepared by applying a π/2 pulse to the qubit
and followed by a swap operation with the phonon mode. From the
measured data shown in Fig. 4a–c, we can reconstruct the measured
state using a maximum likelihood method[18] (see Supplementary
Information). The Wigner functions of the reconstructed states are
presented in Fig. 4d–f. Figure 4g–i shows that the reconstructed parities
agree well with the raw data. The negativity of the Wigner functions
clearly demonstrates the quantum nature of the states. From the recon-
structed density matrices, we find that the fidelities of the prepared
states to the target states are $F_{|1\rangle} = 0.87 \pm 0.01$, $F_{(|0\rangle+|1\rangle)/\sqrt{2}} = 0.94 \pm 0.01$
and $F_{|2\rangle} = 0.78 \pm 0.02$. The infidelity for all three states is dominated by
excess population in the lower-number Fock states, which is an
expected consequence of energy decay during state preparation and
measurement (see Supplementary Information).

These results show that the quantum state of motion in a macroscale
mechanical resonator can be prepared, controlled and fully character-
ized in a circuit QAD device. The demonstration of even more com-
plex quantum states should be possible with further improvements
of the device performance. Currently, the dominant source of loss is
the qubit, and we find that its T_1 is higher when the resonator chip is
either not present or rotated by 180° relative to the qubit chip. This
indicates that the qubit lifetime may be limited by loss due to the AlN,
which could be mitigated by using a different piezoelectric material or
optimizing the device geometry to minimize the electric field in the
AlN that does not contribute to transduction. The current limitations
on the phonon coherence also require further investigation. The energy
loss is probably dominated by surface roughness or imperfections in the
fabricated geometry, whereas additional dephasing could result from
thermal excitations and frequency fluctuations of the detuned qubit[23].
In addition, we can characterize the final flip-chip geometry, such as the
spacing and alignment between the chips, more carefully. The assembly

process can then be modified accordingly, potentially leading to further improvements in the coupling and mode selectivity.

The next generation of devices could give us access to even more sophisticated methods for quantum control of the acoustic resonator. Our ħBAR device can almost reach the strong dispersive regime in which circuit QED systems currently operate, which would allow quantum non-demolition measurements of phonon numbers[24] and the application of more sophisticated techniques for generating arbitrary quantum states of harmonic resonators[17,25]. Furthermore, our qubit-cooling technique already takes advantage of the multimode nature of the acoustic resonator. Future experiments could, for example, demonstrate qubit-mediated interactions between multiple modes and the creation of multipartite entangled states of mechanical motion[11,12]. Recent efforts in improving the efficiency of electromechanical and optomechanical transduction with mechanical resonators could enable conversion of quantum information between the microwave and optical domains[14,26]. Beyond the use of acoustic resonators as resources for quantum information, the creation of increasingly complex quantum states in highly coherent mechanical resonators can provide insight into the question of whether quantum superpositions of massive objects are suppressed owing to mechanisms other than environmental decoherence[4,27]. In addition, the ability to apply quantum control on our large-effective-mass, high-frequency and low-thermal-occupation mechanical system may put new bounds on modifications to quantum mechanics at small length scales[28,29]. These examples suggest that the wide range of quantum acoustics demonstrations that may soon be possible with ħBAR will give rise to new quantum technologies while furthering our understanding of fundamental physics.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

1. O'Connell, A. D. et al. Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703 (2010).
2. Moores, B. A., Sletten, L. R., Viennot, J. J. & Lehnert, K. W. Cavity quantum acoustic device in the multimode strong coupling regime. *Phys. Rev. Lett.* **120**, 227701 (2018).
3. Chu, Y. et al. Quantum acoustics with superconducting qubits. *Science* **358**, 199–202 (2017).
4. Arndt, M. & Hornberger, K. Testing the limits of quantum mechanical superpositions. *Nat. Phys.* **10**, 271–277 (2014).
5. Marshall, W., Simon, C., Penrose, R. & Bouwmeester, D. Towards quantum superpositions of a mirror. *Phys. Rev. Lett.* **91**, 130401 (2003).
6. Leibfried, D. et al. Experimental determination of the motional quantum state of a trapped atom. *Phys. Rev. Lett.* **77**, 4281–4285 (1996).
7. Manenti, R. et al. Circuit quantum acoustodynamics with surface acoustic waves. *Nat. Commun.* **8**, 975 (2017).
8. Riedinger, R. et al. Non-classical correlations between single photons and phonons from a mechanical oscillator. *Nature* **530**, 313–316 (2016).
9. Lee, K. C. et al. Entangling macroscopic diamonds at room temperature. *Science* **334**, 1253–1256 (2011).
10. Safavi-Naeini, A. H. et al. Squeezed light from a silicon micromechanical resonator. *Nature* **500**, 185–189 (2013).
11. Riedinger, R. et al. Remote quantum entanglement between two micromechanical oscillators. *Nature* **556**, 473–477 (2018).
12. Ockeloen-Korppi, C. F. et al. Stabilized entanglement of massive mechanical oscillators. *Nature* **556**, 478–482 (2018).
13. Gustafsson, M. V. et al. Propagating phonons coupled to an artificial atom. *Science* **346**, 207–211 (2014).
14. Andrews, R. W. et al. Bidirectional and efficient conversion between microwave and optical light. *Nat. Phys.* **10**, 321–326 (2014).
15. Schuetz, M. J. A. et al. Universal quantum transducers based on surface acoustic waves. *Phys. Rev. X* **5**, 031031 (2015).
16. Naik, R. K. et al. Random access quantum information processors using multimode circuit quantum electrodynamics. *Nat. Commun.* **8**, 1904 (2017); publisher correction 9, 172 (2018).
17. Leghtas, Z. et al. Hardware-efficient autonomous quantum memory protection. *Phys. Rev. Lett.* **111**, 120501 (2013).
18. Chou, K. S. et al. Deterministic teleportation of a quantum gate between two logical qubits. *Nature* **561**, 368–373 (2018).
19. Satzinger, K. J. et al. Quantum control of surface acoustic-wave phonons. *Nature* https://doi.org/10.1038/s41586-018-0719-5 (2018).
20. Hofheinz, M. et al. Generation of Fock states in a superconducting quantum circuit. *Nature* **454**, 310–314 (2008).
21. Hofheinz, M. et al. Synthesizing arbitrary quantum states in a superconducting resonator. *Nature* **459**, 546–549 (2009).
22. Royer, A. Wigner function as the expectation value of a parity operator. *Phys. Rev. A* **15**, 449–450 (1977).
23. Gambetta, J. et al. Qubit-photon interactions in a cavity: measurement-induced dephasing and number splitting. *Phys. Rev. A* **74**, 042318 (2006).
24. Schuster, D. I. et al. Resolving photon number states in a superconducting circuit. *Nature* **445**, 515–518 (2007).
25. Heeres, R. W. et al. Implementing a universal gate set on a logical qubit encoded in an oscillator. *Nat. Commun.* **8**, 94 (2017).
26. Kharel, P. et al. Ultra-high-Q phononic resonators on-chip at cryogenic temperatures. *APL Photonics* **3**, 066101 (2018).
27. Penrose, R. On gravity's role in quantum state reduction. *Gen. Relativ. Gravit.* **28**, 581–600 (1996).
28. Pikovski, I., Vanner, M. R., Aspelmeyer, M., Kim, M. S. & Brukner, V. Probing planck-scale physics with quantum optics. *Nat. Physics* **8**, 393–397 (2012).
29. Marin, F. et al. Gravitational bar detectors set limits to Planck-scale physics on macroscopic variables. *Nat. Phys.* **9**, 71–73 (2013).

Author contributions Y.C. performed the experiment and analysed the data under the supervision of P.T.R. and R.J.S. Y.C., P.K. and L.F. designed and fabricated the device. P.K. and T.Y. provided experimental suggestions and theory support. Y.C., P.K., P.T.R. and R.J.S. wrote the manuscript with contributions from all authors.

LETTER

Superfluorescence from lead halide perovskite quantum dot superlattices

Gabriele Rainò[1,2,3,5]*, Michael A. Becker[3,4,5], Maryna I. Bodnarchuk[2], Rainer F. Mahrt[3], Maksym V. Kovalenko[1,2]* & Thilo Stöferle[3]*

An ensemble of emitters can behave very differently from its individual constituents when they interact coherently via a common light field. After excitation of such an ensemble, collective coupling can give rise to a many-body quantum phenomenon that results in short, intense bursts of light—so-called superfluorescence[1]. Because this phenomenon requires a fine balance of interactions between the emitters and their decoupling from the environment, together with close identity of the individual emitters, superfluorescence has thus far been observed only in a limited number of systems, such as certain atomic and molecular gases and a few solid-state systems[2–7]. The generation of superfluorescent light in colloidal nanocrystals (which are bright photonic sources practically suited for optoelectronics[8,9]) has been precluded by inhomogeneous emission broadening, low oscillator strength, and fast exciton dephasing. Here we show that caesium lead halide (CsPbX$_3$, X = Cl, Br) perovskite nanocrystals[10–13] that are self-organized into highly ordered three-dimensional superlattices exhibit key signatures of superfluorescence. These are dynamically red-shifted emission with more than 20-fold accelerated radiative decay, extension of the first-order coherence time by more than a factor of four, photon bunching, and delayed emission pulses with Burnham–Chiao ringing behaviour[14] at high excitation density. These mesoscopically extended coherent states could be used to boost the performance of opto-electronic devices[15] and enable entangled multi-photon quantum light sources[16,17].

Spontaneous emission of photons—such as happens in the process of fluorescence that is commonly used in displays and lighting—occurs because of coupling between excited two-level systems (TLS) and the vacuum modes of the electromagnetic field, effectively stimulated by its zero-point fluctuations. In 1954, Dicke predicted[18] that an ensemble of N identical TLS confined in a volume smaller than about λ^3 (where λ is the corresponding emission wavelength of the TLS) can exhibit coherent and cooperative spontaneous emission. This so-called super-radiant emission results from the coherent coupling between individual TLS through the common vacuum modes, effectively leading to a single giant emitting dipole from all participating TLS. Superradiant emission has been observed in distinctly different physical systems, such as molecular aggregates and crystals[19], nitrogen vacancy centres in diamond[20] and epitaxially grown quantum dots[21] (QDs). In the case when the excited TLS are initially fully uncorrelated, the coherence can be established only through spontaneously triggered correlations due to quantum fluctuations rather than by coherent excitation. When this occurs, a so-called superfluorescence (SF) pulse is emitted[1] (Fig. 1, illustrated for the present study). Both superradiant emission and coherent SF bursts are characterized by an accelerated radiative decay time $\tau_{SF} \propto \tau_{SE}/N$, where the exponential decay time τ_{SE} of spontaneous emission from the uncoupled TLS is shortened by the number of coupled emitters N. In addition, SF exhibits the following fundamental signatures, the magnitudes of which are also dependent on the excitation density: (i) a delay or build-up time $\tau_D \propto \log(N)/N$ during which the emitters couple and phase-synchronize to each other, and which

corresponds to the time delay between the excitation and onset of the cooperative emission (Fig. 1); and (ii) coherent Rabi-type oscillations in the time domain due to the strong light–matter interaction, known as Burnham–Chiao ringing[14,22].

Superfluorescence was first observed in a dense gas of hydrogen fluoride[2], and then in a limited number of solid-state systems, such as CuCl nanocrystals (NCs) formed in a NaCl matrix[4], KCl crystals doped with peroxide anions[3] (O$_2^-$), and some select semiconductor crystals (ZnTe and InGaAs/GaAs multi-quantum wells)[5,6]. Practical implementation of such an enhanced radiative property is a persistent challenge. Besides stringent requirements for the emissive material (for example, high oscillator strength, small inhomogeneous line-broadening, small exciton dephasing), equally important are structural, optical and device engineerability. Colloidal semiconductor NCs, also known as colloidal QDs, could fill this gap as they are structurally and optically versatile, and highly suited for the entire visible spectral range. Although they are actively pursued for photonic applications[8,9,23], they have not been reported to exhibit SF.

Here we use colloidal NCs of caesium lead halide perovskites (CsPbX$_3$, X = Cl, Br) that can be synthesized with narrow size dispersion and are known to exhibit moderate quantum confinement effects, resulting in narrow-band emission combined with exceptionally large oscillator strength from a bright triplet state[10,11,24]. In order to foster cooperative behaviour, we employ structurally well-defined, long-range-ordered, and densely packed arrays of such NCs, known as superlattices, produced by means of solvent-drying-induced spontaneous assembly[25–28]. Similarly, regular arrays of II–VI semiconductor NCs have been used to obtain collective effects in the electronic domain, that is, band-like transport[15]. Figure 2a outlines superlattice formation (see also Methods), using a solution of highly monodispersed CsPbBr$_3$ NCs with a mean size of 9.5 nm and size standard deviation of less
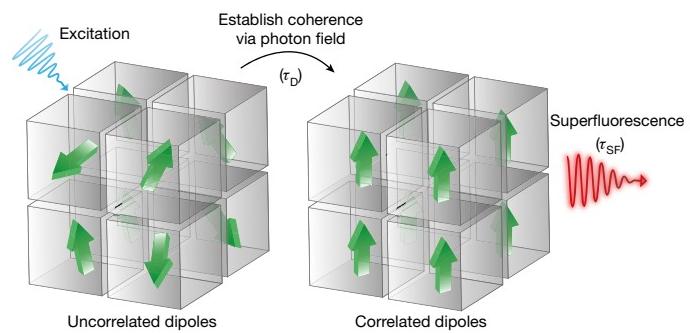
Fig. 1 | Schematic of the build-up process of SF. An initially uncorrelated ensemble of TLS (randomly oriented green arrows) is excited by a light pulse (blue arrow, top left). After time τ_D their phases are synchronized (aligned green arrows) such that they cooperatively emit a SF light pulse (red arrow at right) with a characteristic decay time τ_{SF}. Grey cubes represent long-range-ordered self-assembled superlattices.

[1]Department of Chemistry and Applied Bioscience, Institute of Inorganic Chemistry, ETH Zurich, Zurich, Switzerland. [2]Laboratory of Thin Films and Photovoltaics, Empa — Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland. [3]IBM Research — Zurich, Rüschlikon, Switzerland. [4]Optical Materials Engineering Laboratory, ETH Zurich, Zurich, Switzerland. [5]These authors contributed equally: G. Rainò; M. A. Becker. *e-mail: rainog@ethz.ch; mvkovalenko@ethz.ch; tof@zurich.ibm.com
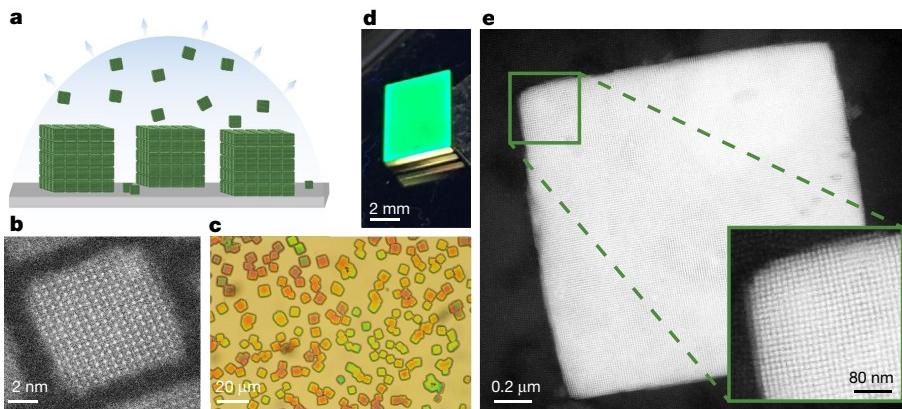
Fig. 2 | Formation of CsPbX₃ (X = Cl, Br) NC superlattices by drying-mediated self-assembly. a, Illustration of the assembly process: a controlled evaporation (indicated by the pale blue arrows) of the solvent leads to the formation of micrometre-size cuboidal superlattices. **b**, High-resolution image (obtained using high-angle annular dark-field scanning transmission electron microscopy, HAADF–STEM) of a single CsPbBr₃ NC. **c**, Optical microscope image and **d**, photograph (under ultraviolet light) of a layer of micrometre-sized, three-dimensional, cuboidal-shaped NC superlattices. **e**, HAADF–STEM image of a single superlattice of CsPbBr₃ NCs. The cubic shape of the individual perovskite NC building blocks is translated into the symmetry of the superlattice (simple cubic packing). Inset, magnified view of the boxed area, showing the individual NCs.

than 5% (Extended Data Fig. 1). In the self-assembly process, cuboidal individual superlattice domains are formed (that is, supercrystals), each consisting of up to several million NCs. Optical microscopy (Fig. 2c) reveals superlattices with a lateral size of up to 5 μm, randomly distributed in a uniform film on a 5 mm × 7 mm sample (Fig. 2d). Transmission electron microscopy confirms that highly ordered super-lattices consist of well-separated individual NCs (Fig. 2e and Extended Data Fig. 2). More details of the self-assembly process are reported in the Methods section.

Figure 3a displays the photoluminescence (PL) spectrum of a single CsPbBr₃ superlattice (excited at 3.06 eV) exhibiting two emission peaks. This and all other optical measurements were performed at a temperature of 6 K in vacuum or in a helium atmosphere (see Methods for details). The high-energy emission peak coincides with the centre energy of PL from a disordered dense film of CsPbBr₃ NCs (in a glassy state) and is therefore assigned to uncoupled QDs. In addition, a narrow, red-shifted emission peak appears in PL from superlattices, which we assign to the emission of coupled QDs, which is best fitted with a Lorentzian (full-width at half-maximum, FWHM$_{coupled}$ = 11 meV). The peaks from the uncoupled QDs in a superlattice and in the glassy films are best fitted with a Gaussian, as expected for disordered ensembles. The width of the uncoupled QD

emission (FWHM$_{uncoupled}$ = 55 meV) is slightly broader than that of the amorphous film (FWHM$_{amorphous}$ = 35 meV), which can be explained by assuming that more 'identical QDs' within the superlattice are now forming the peak from coupled QDs while the remaining uncoupled ones appear more disordered than the inhomogeneous energy distri-bution of the primary QD material. We can exclude the possibility that the red-shifted feature, which is about 70 meV lower in energy than the uncoupled QD emission, originates from emission from trions, bi-excitons or multi-excitons because the energy shifts of these last three species are reportedly[11,12] 10−20 meV, and these would also be observable in the disordered ensemble. The number and interaction strength of coupled QDs determine the magnitude of the energy shift. Statistics from 10 superlattices from different samples give an average static red-shift of (64 ± 6) meV, average FWHM$_{coupled}$ = (15 ± 4) meV and average FWHM$_{uncoupled}$ = (49 ± 21) meV. In most superlattices, we observe a substructure in this red-shifted emission band, which we attribute to the presence of several slightly different independent domains within the same individual superlattice.

A central feature of cooperative emission is the modification of the radiative lifetime[18], as demonstrated experimentally in several quantum emitters[6,20,21]. In time-resolved PL decay measurements at a very low excitation fluence (5 nJ cm⁻²), we do not observe a significant
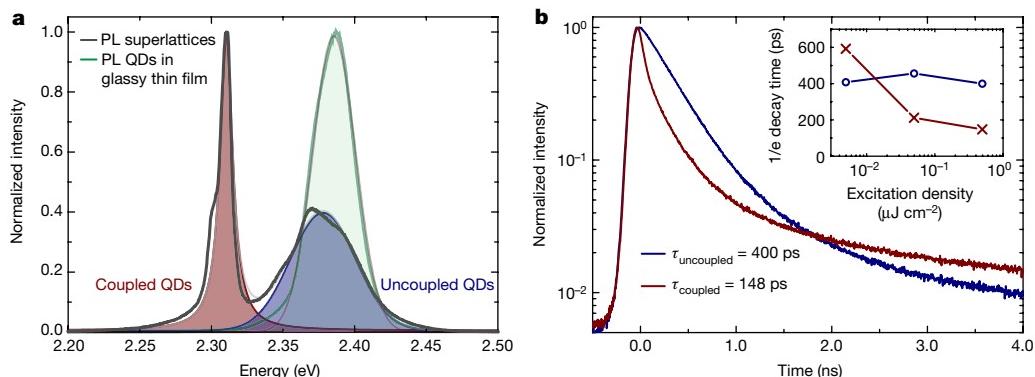
Fig. 3 | Optical properties of CsPbBr₃ QD superlattices. a, PL spectrum of a single CsPbBr₃ superlattice (black solid line). The high-energy band is assigned to the emission of uncoupled QDs. The low-energy band is the result of the emission of coupled QDs and is not present in glassy films of NCs (green solid line). The shaded areas are fits to the data (see main text). **b**, Time-resolved PL decay of the two emission bands at 500 nJ cm⁻² excitation fluence after applying suitable spectral filters to separate the

two components (blue curve, from uncoupled QDs; dark-red curve, from coupled QDs). The 1/e-decay time of the two bands ($\tau_{coupled}$ and $\tau_{uncoupled}$, respectively) are also indicated. With increasing excitation fluence, the decay from the coherently coupled QDs is substantially faster than from the uncoupled ones. Inset, power-dependence of the 1/e-decay times of both components.

modification of the decay of the coupled QD emission compared to that of the uncoupled QD emission (Fig. 3b inset). The absence of accelerated emission at vanishing excitation fluence and the presence of the red-shifted feature in PL excitation scans (Extended Data Fig. 3) corroborate that the static red-shift of ~70 meV originates from incoherent coupling of the QDs in the ground state, similar to that found in various molecular aggregates[19]. At a slightly higher excitation fluence (500 nJ cm^{-2} per pulse), we already observe an accelerated PL decay of the coupled QD emission peak in comparison to the PL decay of uncoupled QDs, with $1/e$ decay times of $\tau_{SF} = 148$ ps and $\tau_{QD} = 400$ ps, respectively (Fig. 3b). In contrast to the predominantly mono-exponential decay of the uncoupled QDs, this SF emission decay is well approximated by a stretched exponential[29] (see Methods section), because the number of excited coupled emitters, and therefore the emission acceleration, varies during the decay. Furthermore, in contrast to the uncoupled QDs, the SF decay time is strongly dependent on excitation power (Fig. 3b inset): this is because it scales with the coupling strength among the QDs, which is given by the intensity of the common light-field that effectively corresponds to a change in the number of coherently coupled QDs. When the spectrally and temporally integrated emission is fitted with a power law, we obtain an exponent of 1 (Extended Data Fig. 4), indicating that excitation-density-dependent non-radiative decay channels (for example, Auger recombination) are absent. Notably, no threshold behaviour as occurs for amplified spontaneous emission is observed.

The cooperative emission process strongly influences the coherence of the emitted light. First-order correlation measurements of each of the two emission peaks by means of a Michelson interferometer allow us to monitor the interference pattern and therefore the phase coherence time (Fig. 4a). The emission band of the uncoupled QDs exhibits a coherence time of 38 fs, best fitted with a Gaussian decay (Fig. 4a, upper graph), typical of incoherent (thermal) light sources. The emission from the coherently coupled QDs (Fig. 4a, lower graph) exhibits a much longer coherence time with an exponential decay of 140 fs. For some superlattices, a Gaussian decay is observed (Extended Data Fig. 5a), which might be attributed to number fluctuations within the coherent SF state[30].

Second-order coherence of the emitted light is evinced by the statistics of the photon arrival time on a detector[31]. Typical coherent light, as from a laser, shows a random (Poissonian) distribution of photon arrival times, whereas a single TLS exhibits photon antibunching (a sub-Poissonian distribution). In contrast, the cooperative emission from coupled QDs leads to coherent multi-photon emission bursts. Figure 4b reports the second-order correlation function, $g^{(2)}(\tau) = \langle I(t)I(t+\tau)\rangle/\langle I(t)\rangle^2$, for both PL emission bands, where $I(t)$ is the signal intensity at time t and $\langle I(t)\rangle$ is its statistical average. For the uncoupled QD emission (Fig. 4b, upper graph), the plot is flat ($g^{(2)}(\tau) = 1$) because the experimental temporal resolution (40 ps) is insufficient to resolve the expected thermal bunching. The SF emission band, however, shows pronounced photon bunching (Fig. 4b, lower graph) because the coherent coupling leads to the correlated emission of multiple photons within a short time interval. Photon bunching is only observable in superlattices with one or a few SF domains (that is, where no substructure is visible in the red-shifted emission band) because spectrally overlapping uncorrelated aggregated domains within the same superlattice reduce the bunching peak's visibility, as predicted by theory[32]. Yet it is a robust effect that is observed with pulsed excitation and also for mixed-halide (CsPbBr$_2$Cl, emitting at higher energies) QD superlattices (see Extended Data Figs. 5b and 6b, respectively). Remarkably, some superlattices with supposedly well-isolated coherently coupled QDs exhibit $g^{(2)}(0) > 2$ (Fig. 4 inset), similarly to superthermal emission[31]. The exponential decay time of the second-order correlation is of the order of the radiative decay time of the SF emission for low excitation densities ($\tau_{g^{(2)}} = 224$ ps).

SF emission exhibits very distinct characteristics in the time domain under strong driving conditions. Figure 5a shows a streak camera image acquired at an excitation fluence of 600 μJ cm^{-2}, where a finite
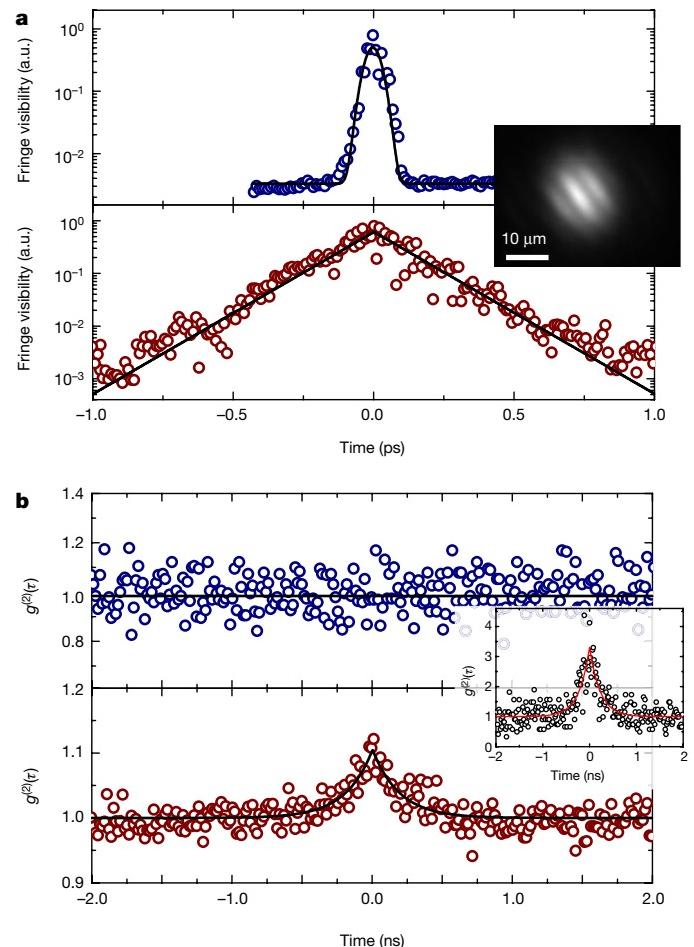
Fig. 4 | First- and second-order coherence properties of CsPbBr$_3$ QD superlattices. a, First-order correlation of the two emission bands as obtained from the interference fringe visibility using a Michelson interferometer. The high-energy band of the uncoupled QDs has a very short phase coherence time (<40 fs, upper graph, blue symbols), whereas the red-shifted band from the coupled QDs is characterized by much longer phase coherence (140 fs, lower graph, dark-red symbols). The solid lines are fits to the data (see main text). a.u., arbitrary units. Inset, an example of the real-space interferogram. **b**, Second-order correlation function, $g^{(2)}(\tau)$, obtained with a Hanbury–Brown and Twiss set-up in start–stop configuration. For the high-energy band (upper graph, blue symbols), a flat profile with $g^{(2)}(\tau) = 1$ is observed. The red-shifted emission band (lower graph, dark-red symbols) from the SF emission displays a pronounced bunching peak, characteristic of the correlated emission during a photon burst. The data are fitted to the function $g^{(2)}(\tau) = 1 + A\exp(-|\tau|/\tau_c)$ (solid lines), where A is an amplitude prefactor, and τ_c the characteristic decay time of the second order coherence. Inset, an example of superbunching with $g^{(2)}(0) > 2$ from a single superlattice. Black open circles are experimental data while the red curve represents the best fit to the function described above.

rise time and subsequent oscillation of the emission are observed in addition to a much shortened radiative decay. Quantitative analysis of spectrally integrated PL decay traces for various excitation power densities is shown in Fig. 5b (for details see Methods). As excitation fluence is increased, the decay time shortens to 14 ps (Fig. 5c, upper panel). From this shortening, which is an order of magnitude stronger than that reported for the collective emission from other QD systems[4,21], we can estimate the average number of coherently coupled QDs to be $N \approx 28$. This is only an effective value and is a conservative estimate, because the energetic disorder of the QD emission energies (FWHM$_{coupled}$ = 11 meV) still substantially exceeds the emission peak width of individual QDs (typically[11] FWHM ≈ 1 meV) and thereby effectively reduces the coherent coupling[33]. The SF emission undergoes
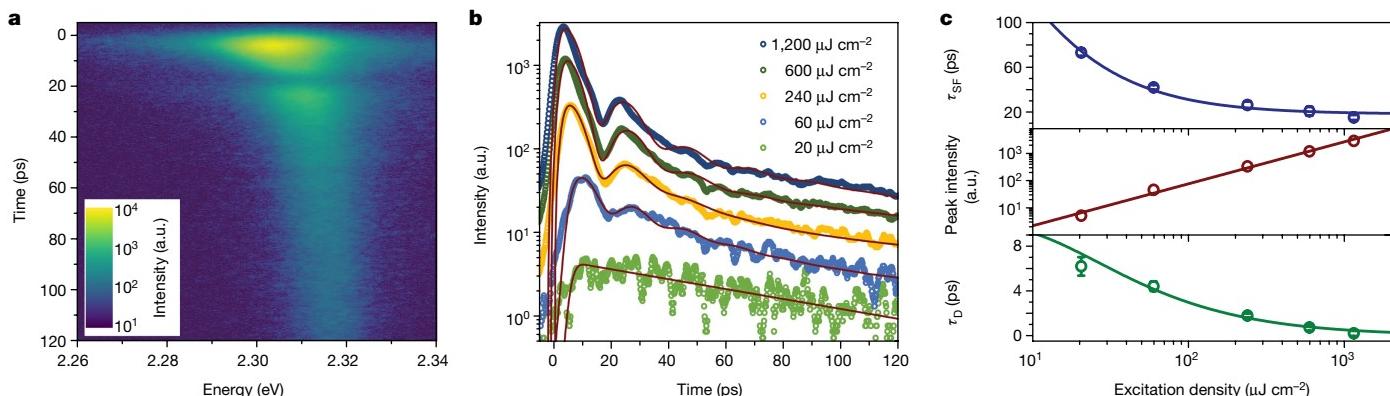
Fig. 5 | Burnham–Chiao ringing behaviour of CsPbBr₃ QD superlattices. **a**, Streak camera image of SF dynamics obtained with a high excitation density of 600 μJ cm⁻². See Methods for details. **b**, Extracted time-resolved emission intensity traces for five different excitation powers (see key). Solid red lines are weighted best-fits to a model that employs a bi-exponential decay function with damped oscillations. **c**, Top, effective SF decay (τ_{SF}, blue circles) as a function of the excitation power density fitted according to the SF model (solid blue line). Middle, dark-red circles represent the peak SF emission intensity that increases superlinearly with excitation power, corresponding to a power-law dependence with an exponent $\alpha = 1.5 \pm 0.1$ (solid dark-red line). Bottom, the extracted delay time τ_D (green circles) decreases at high excitation power due to the increased interaction among the emitters. The green solid line is the best fit according the model described in the Methods section. The error bars represent the parameters' fit uncertainty.

a dynamical red-shift of up to 15 meV owing to renormalization of the emission energy from the coherent coupling[7], which decreases in the course of the decay as the number of excited dipoles reduces (Extended Data Fig. 7). The peak intensity increases superlinearly over three orders of magnitude (Fig. 5c, middle panel) according to a power-law dependence with an exponent of $\alpha = 1.5 \pm 0.1$, deviating from the theoretically expected[7] value of $\alpha = 2$, presumably owing to saturation effects[5]. Nevertheless, no substantial quenching of the emission for high excitation fluences was found, verifying that the decay remains essentially radiative (Extended Data Fig. 4d). Furthermore, a shortening of the SF build-up time (τ_D), after which the photon burst is emitted, is observed (Fig. 5c, bottom panel). This characteristic of SF is a consequence of the time it takes for the individual dipoles to become phase-locked and scales with the number N of excited coupled QDs according to $\tau_D \propto \log(N)/N$ (see Methods section).

As SF crucially depends on low decoherence and low inhomogeneous energy variation, it should be noted that SF coupling is strongly affected by the environment around the QDs (for example, the number of free ligands, which are the molecules used to passivate the NC surface), by the superlattice assembly, and by the quality of the QDs themselves. Thus, while a large fraction of the superlattices display the red-shifted emission from coupled QDs, the amount of photon-bunching and Burnham–Chiao ringing varied from superlattice to superlattice. However, experiments employing different batches of NCs and super-lattice assemblies of CsPbBr₃ and CsPbBr₂Cl NCs (see Extended Data Figs. 6–9) were consistently reproducible, but further optimization of the synthesis and assembly is likely to improve the yield of SF domains. It is important to note that experiments on control samples with diluted, uncoupled QDs under similar excitation conditions do not show any of the above-mentioned signatures of SF (Extended Data Fig. 10), proving that the observed peculiar emission characteristics of QD superlattices arise from a genuine multi-particle effect.

Our measurements reveal that coherent SF coupling can be achieved in long-range-ordered self-assembled superlattices of fully inorganic CsPbX₃ perovskite NCs, resulting in strong emission bursts. Colloidal NCs and their assemblies have proven to be excellent building blocks for a large variety of opto-electronic devices, and these cooperative effects now allow modification of the opto-electronic properties beyond what is possible at the individual QD level with chemical engineering approaches. This opens up new opportunities for high-brightness and multi-photon quantum light sources, and could enable the exploitation of cooperative effects for long-range quantum transport and ultra-narrow tunable lasers.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0683-0.

1. Bonifacio, R. & Lugiato, L. A. Cooperative radiation processes in two-level systems: superfluorescence. *Phys. Rev. A* **11**, 1507–1521 (1975).
2. Skribanowitz, N., Herman, I. P., MacGillivray, J. C. & Feld, M. S. Observation of Dicke superradiance in optically pumped HF gas. *Phys. Rev. Lett.* **30**, 309–312 (1973).
3. Malcuit, M. S., Maki, J. J., Simkin, D. J. & Boyd, R. W. Transition from superfluorescence to amplified spontaneous emission. *Phys. Rev. Lett.* **59**, 1189–1192 (1987).
4. Miyajima, K., Kagotani, Y., Saito, S., Ashida, M. & Itoh, T. Superfluorescent pulsed emission from biexcitons in an ensemble of semiconductor quantum dots. *J. Phys. Condens. Matter* **21**, 195802 (2009).
5. Dai, D. C. & Monkman, A. P. Observation of superfluorescence from a quantum ensemble of coherent excitons in a ZnTe crystal: evidence for spontaneous Bose-Einstein condensation of excitons. *Phys. Rev. B* **84**, 115206 (2011).
6. Noe, G. T. II et al. Giant superfluorescent bursts from a semiconductor magneto-plasma. *Nat. Phys.* **8**, 219–224 (2012).
7. Cong, K. et al. Dicke superradiance in solids. *J. Opt. Soc. Am. B* **33**, C80–C101 (2016).
8. Talapin, D. V., Lee, J.-S., Kovalenko, M. V. & Shevchenko, E. V. Prospects of colloidal nanocrystals for electronic and optoelectronic applications. *Chem. Rev.* **110**, 389–458 (2010).
9. Kovalenko, M. V. et al. Prospects of nanoscience with nanocrystals. *ACS Nano* **9**, 1012–1057 (2015).
10. Protesescu, L. et al. Nanocrystals of cesium lead halide perovskites (CsPbX₃, X=Cl, Br, and I): novel optoelectronic materials showing bright emission with wide color gamut. *Nano Lett.* **15**, 3692–3696 (2015).
11. Rainò, G. et al. Single cesium lead halide perovskite nanocrystals at low temperature: fast single-photon emission, reduced blinking, and exciton fine structure. *ACS Nano* **10**, 2485–2490 (2016).
12. Fu, M. et al. Neutral and charged exciton fine structure in single lead halide perovskite nanocrystals revealed by magneto-optical spectroscopy. *Nano Lett.* **17**, 2895–2901 (2017).
13. Kovalenko, M. V., Protesescu, L. & Bodnarchuk, M. I. Properties and potential optoelectronic applications of lead halide perovskite nanocrystals. *Science* **358**, 745–750 (2017).
14. Burnham, D. C. & Chiao, R. Y. Coherent resonance fluorescence excited by short light pulses. *Phys. Rev.* **188**, 667–675 (1969).
15. Kagan, C. R. & Murray, C. B. Charge transport in strongly coupled quantum dot solids. *Nat. Nanotechnol.* **10**, 1013–1026 (2015).
16. Afek, I., Ambar, O. & Silberberg, Y. High-NOON states by mixing quantum and classical light. *Science* **328**, 879–881 (2010).
17. Muñoz, C. S. et al. Emitters of N-photon bundles. *Nat. Photon.* **8**, 550–555 (2014).
18. Dicke, R. H. Coherence in spontaneous radiation processes. *Phys. Rev.* **93**, 99–110 (1954).
19. Spano, F. C. The spectral signatures of Frenkel polarons in H- and J-aggregates. *Acc. Chem. Res.* **43**, 429–439 (2010).

20. Bradac, C. et al. Room-temperature spontaneous superradiance from single diamond nanocrystals. *Nat. Commun.* **8**, 1205 (2017).
21. Scheibner, M. et al. Superradiance of quantum dots. *Nat. Phys.* **3**, 106–110 (2007).
22. Heinzen, D. J., Thomas, J. E. & Feld, M. S. Coherent ringing in superfluorescence. *Phys. Rev. Lett.* **54**, 677–680 (1985).
23. Shirasaki, Y., Supran, G. J., Bawendi, M. G. & Bulović, V. Emergence of colloidal quantum-dot light-emitting technologies. *Nat. Photon.* **7**, 13–23 (2013).
24. Becker, M. A. et al. Bright triplet excitons in caesium lead halide perovskites. *Nature* **553**, 189–193 (2018).
25. Boles, M. A., Engel, M. & Talapin, D. V. Self-assembly of colloidal nanocrystals: from intricate structures to functional materials. *Chem. Rev.* **116**, 11220–11289 (2016).
26. Murray, C. B., Kagan, C. R. & Bawendi, M. G. Synthesis and characterization of monodisperse nanocrystals and close-packed nanocrystal assemblies. *Annu. Rev. Mater. Sci.* **30**, 545–610 (2000).
27. Geuchies, J. J. et al. In situ study of the formation mechanism of two-dimensional superlattices from PbSe nanocrystals. *Nat. Mater.* **15**, 1248–1254 (2016).
28. Nagaoka, Y. et al. Nanocube superlattices of cesium lead bromide perovskites and pressure-induced phase transformations at atomic and mesoscale levels. *Adv. Mater.* **29**, 1606666 (2017).
29. Temnov, V. V. & Woggon, U. Superradiance and subradiance in an inhomogeneously broadened ensemble of two-level systems coupled to a low-Q cavity. *Phys. Rev. Lett.* **95**, 243602 (2005).
30. Whittaker, D. M. & Eastham, P. R. Coherence properties of the microcavity polariton condensate. *Europhys. Lett.* **87**, 27002 (2009).
31. Jahnke, F. et al. Giant photon bunching, superradiant pulse emission and excitation trapping in quantum-dot nanolasers. *Nat. Commun.* **7**, 11540 (2016).
32. Temnov, V. V. & Woggon, U. Photon statistics in the cooperative spontaneous emission. *Opt. Express* **17**, 5774 (2009).
33. Ishikawa, A., Miyajima, K., Ashida, M., Itoh, T. & Ishihara, H. Theory of superfluorescence in highly inhomogeneous quantum systems. *J. Phys. Soc. Jpn.* **85**, 034703 (2016).

Author contributions The work originated from continuing interactions between G.R., M.V.K., R.F.M. and T.S. M.I.B. prepared the samples and performed their structural characterization. G.R., M.A.B. and T.S. performed the optical experiments, and interpreted the data with input from R.F.M. G.R. and M.A.B. wrote the manuscript with input from all the co-authors. R.F.M., M.V.K. and T.S. supervised the work.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0683-0.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to G.R., M.V.K. and T.S.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Synthesis of CsPbBr₃ nanocrystals. In a 25 ml three-necked flask, $PbBr_2$ (69 mg, 0.188 mmol, Aldrich, 99%) was suspended in octadecene (5 ml), dried at 100 °C for 30 min, and mixed with oleic acid (0.5 ml, vacuum-dried at 100 °C) and oleylamine (0.5 ml vacuum-dried at 100 °C). When the $PbBr_2$ had dissolved, the reaction mixture was heated up to 180 °C and preheated caesium oleate in octadecene (0.4 ml, 0.125 M) was injected. The reaction mixture was cooled immediately with an ice bath to room temperature.

Synthesis of CsPbBr₂Cl nanocrystals. In a 25 ml three-necked flask, $PbBr_2$ (45 mg, 0.12 mmol, Aldrich, 99%), $PbCl_2$ (18 mg, 0.064 mmol, ABCR) and 1 ml trioctylphosphine (Strem, 97%) was suspended in octadecene (5 ml), dried at 100 °C for 30 min, and mixed with oleic acid (0.5 ml, vacuum-dried at 100 °C) and oleylamine (0.5 ml vacuum-dried at 100 °C). When the $PbCl_2$ and $PbBr_2$ had dissolved, the reaction mixture was heated up to 180 °C and preheated caesium oleate in octadecene (0.4 ml, 0.125 M) was injected. The reaction mixture was cooled immediately with an ice bath to room temperature.

Purification and size-selection of CsPbX₃ (X = Cl, Br) nanocrystals. A critical factor for self-assembly of cubic-shaped CsPbX₃ NCs is to start with an initially high level of monodispersity. The crude solution was centrifuged at 12,100 revolutions per minute for 5 min, following which the supernatant was discarded, and the precipitate was dissolved in 300 μl hexane. The hexane solution was centrifuged again and the precipitate was discarded. The supernatant was diluted twice and used for further purification. Subsequently, two methods of purification of the NCs were applied: (a) 50 μl hexane, 0.6 μl oleic acid and 0.6 μl oleylamine were added to 50 μl NCs in hexane. The colloid was destabilized by adding 50 μl acetone, followed by centrifuging and dispersing the NCs in 300 μl toluene. This solution was used further for the preparation of the 3D superlattices. (b) 50 μl hexane and 100 μl toluene were added to 50 μl NCs in hexane. The colloid was destabilized by adding 50 μl acetonitrile, followed by centrifuging and dispersing the NCs in 300 μl toluene. This solution was used further for the preparation of the 3D superlattices.

Preparation of 3D superlattices. CsPbX₃ NC superlattices were prepared on glass or on 5 mm × 7 mm silicon substrates. Shortly before the self-assembly process, the silicon substrate was dipped into 4% solution of HF in water for 1 min, followed by washing with water. In a typical assembly process, the substrate was placed in a 10 mm × 10 mm × 10 mm Teflon well and 10 μl of purified NCs in toluene were spread onto the substrate. The well was covered with a glass slide and the toluene was then allowed to evaporate slowly. 3D superlattices of CsPbBr₃ NCs were formed upon complete evaporation of the toluene. Typical lateral dimensions of individual superlattices ranged from 1 μm to 10 μm, wherein some of them are arranged into clusters of several superlattices and others remain spatially well-isolated so that PL measurements can be performed on an individual superlattice. Greater purification or greater polydispersity of NCs led to disordered or 2D assemblies (glassy films). Furthermore, the formation of NC superlattices can serve to further narrow the size distribution and shape uniformity within the ensemble (with smaller or larger NCs being repelled from the NC domain), especially in the case of simple cubic packing of cubes, which is particularly intolerant of size and shape variations.

Optical spectroscopy. All measurements were performed at cryogenic temperatures (6 K). For PL, time-resolved PL, and second-order photon-correlation measurements on single superlattices, the sample was mounted in an evacuated liquid-helium flow cryostat on an *xyz* positioning stage and excited with a fibre-coupled excitation laser at an energy of 3.06 eV, either in continuous wave mode or pulsed mode with a 40 MHz repetition rate (pulse duration 50 ps). The excitation laser output was filtered with a short-pass filter and directed towards the long-working-distance 100× microscope objective (numerical aperture NA = 0.7) by a dichroic beam splitter, resulting in a nearly Gaussian-shaped excitation spot with a $1/e^2$ radius of 1.4 μm. The emission was collected via the same microscope objective and filtered using a tunable bandpass filter. For PL measurements, the collected light was then dispersed by a 300 lines per mm grating inside a 750 mm monochromator and detected by an EMCCD camera. For measurements of the PL decay, we filtered the emission with a tunable band-pass (FWHM = 15 nm) and recorded the decay with an avalanche photodiode single-photon detector with a time resolution of 30 ps connected to a time-correlated single-photon-counting system. The photon correlation was recorded using a similar set-up with two detectors in a Hanbury–Brown and Twiss configuration.

To record streak camera images and first-order coherence measurements, we excited the sample, which was mounted in an exchange-gas cryostat at 6 K, with a frequency-doubled regenerative amplifier seeded with a mode-locked Ti:sapphire laser with a pulse duration of 100–200 fs and a repetition rate of 1 kHz at 3.1 eV. For both excitation and detection, we used an 80 mm lens (NA = 0.013 after iris), resulting in an oval excitation spot area of 20 μm × 40 μm. The recorded PL was dispersed by a grating with 150 lines per mm in a 300 mm spectrograph and detected with a streak camera with a nominal time resolution of 2 ps and an instrument response function FWHM of 4 ps (see Extended Data Fig. 10). First-order coherence measurements were performed using a Michelson interferometer. Here a non-polarizing beam splitter is used to split and recombine the light in the two interferometer arms, with one arm including a retroreflector on a delay stage with 100 nm step resolution. A tunable band-pass filter is applied to select the emission from either the coupled or the uncoupled QDs. The interferogram was recorded as real-space images of the recombined and focused detection beams on camera.

Optical properties of superfluorescence, superadiance and subradiance. As shown in Fig. 3b, we observed that the PL decay of the SF state is initially very fast and cannot be described with a single exponential because the decay rate Γ is dependent on the number of excited TLS, $\Gamma(N) \propto N$, and therefore decreases during the decay. Consequently, the SF decay rate should converge towards the decay rate of the uncoupled nanocrystals. However, we observe that the SF decay trace crosses the bi-exponential PL decay of the uncoupled QDs after 97% of the photons are emitted due to long decay components. These long decay components might originate from coupled QDs where the individual dipoles are out of phase and interfere destructively, a phenomenon known as subradiance[29,34]. In ensembles with inhomogeneously broadened PL, SF and subradiant states can coexist, and we find good agreement of the predicted excited state population with the measured PL decay[35].

An out-of-phase coupling amongst the QDs is expected to result in a higher photon energy of the subradiant state compared to the SF state. In Extended Data Figs. 7 and 9, we provide an analysis of the dynamical energy shift observed at high excitation power density for CsPbBr₃ and CsPbBr₂Cl QD superlattices, respectively. Examples of emission spectra at different times are reported in Extended Data Figs. 7a and 9a for the respective QD halide compositions. In Extended Data Figs. 7b and 9b, we plot the fitted centre photon energy of time-sliced PL spectra (2 ps bin) as a function of the fitted peak area (that is, the time-dependent emission intensity), as obtained from excitation-power-dependent streak camera images, again for both CsPbBr₃ and CsPbBr₂Cl QD superlattices. This effectively shows the energetic shift of the SF state as a function of its occupation, with the different curves representing different initial excitation powers. The green arrows indicate the time sequence of the individual analysed spectral traces. By increasing the excitation power, we observe that the initial dynamical red-shift is largest for the highest excitation power, as is expected from its relationship to the number of excited coupled QDs. Hence, when the number of excited coupled QDs decreases during the decay process, the emission energy blue-shifts to higher energy, as can be seen in Extended Data Figs. 7c and 9c where the fitted centre photon energy is plotted as a function of time. We observe the most pronounced energetic blue-shift for the highest excitation power, resulting in a final emission with a photon energy that has been boosted incrementally more in comparison to the blue-shift for low excitation power, which is another indication of the presence of subradiant states that emit at higher energies. For high excitation power, the SF state becomes depopulated much faster since more QDs are coupled simultaneously. Then, at long timescales after the initial decay, the percentage of subradiant states becomes dominant, resulting in a blue-shift of the PL emission.

Superfluorescence fit model. SF decay traces as in Fig. 3b cannot be fitted well with mono- or bi-exponential functions because the decay rate is proportional to the number of excited coupled QDs, $\Gamma(N) \propto N$, which also decays over time. Furthermore, the resulting characteristic decay follows neither a stretched-exponential nor a power-law dependence[36] exactly, whereas the PL decay of the uncoupled QDs is well described by a bi-exponential behaviour, where the initial fast decay $\tau_{QD} = 349.8 \pm 0.4$ ps accounts for over 96% the total emitted photons. Nevertheless, we found that the best approximate fit to the SF decay trace as a function of time t is the Kohlrausch stretched-exponential decay model $f(t)/f(0) = \exp[-(\Gamma_{streched}t)^{\beta}]$, where $f(0)$ is the initial intensity at $t = 0$, $\Gamma_{streched}$ is the average decay time and $\beta \in [0, 1]$ is the stretch parameter, which represents the distribution of decay rates[37]. Using this model to fit the SF decay curve, we obtain an average decay time $\tau_{stretched} = 40.4 \pm 0.5$ ps and a stretch parameter $\beta = 0.457 \pm 0.002$.

At a high excitation density, as shown in Fig. 5b for CsPbBr₃ QD superlattices and in Extended Data Fig. 8b for CsPbBr₂Cl QD superlattices, we observe oscillations in the decay. To model the SF decay with this characteristic ringing behaviour, we used a decay model consisting of a bi-exponential decay that is multiplied by a damped oscillating term $1 + B\exp(-\gamma_{Damp}t)\cos(\omega t + \phi_0)$, where B is the initial amplitude of the oscillation, ω is its angular frequency, ϕ_0 is its initial phase and γ_{Damp} is its damping constant. Furthermore, for the rising edge of the emitted pulse, which is emitted after build-up time τ_D, we take into account a Gaussian rise term $\exp\{-[(t - \tau_D)/\tau_{rise}]^2\}$, such that the complete fit function is given by[38]:

$$f(t) = \sum_{n=1,2} A_n \, e^{\frac{\tau_{rise}^2}{4\tau_n^2} - \frac{t-\tau_D}{\tau_n}}$$

$$\times \left\{ \frac{1}{2}[1 + Be^{-\gamma_{Damp}(t-\tau_D)}\cos(\omega(t-\tau_D) + \phi_0)] \left[1 + \mathrm{erf}\left(\frac{t-\tau_D}{\tau_{rise}} - \frac{\tau_{rise}}{2\tau_n}\right)\right] \right\}$$

Here, A_n are the amplitudes of the exponential decay with the corresponding decay time constants, τ_n. Both the fast decay time and the long decay time component (Extended Data Fig. 7d for CsPbBr$_3$ QD superlattices and Extended Data Fig. 9d for CsPbBr$_2$Cl QD superlattices) decrease upon increasing the excitation density, whereas the rise time, $\tau_{rise} = 2.4 \pm 0.3$ ps for CsPbBr$_3$ QD superlattices ($\tau_{rise} = 3.4 \pm 1.0$ ps for CsPbBr$_2$Cl QD superlattices), stays approximately constant (probably clamped by the time resolution of the set-up). In the upper panel of Fig. 5c, we plot the power-dependent effective decay time $\tau_{SF} = (A_1\tau_1 + A_2\tau_2)/(A_1 + A_2)$ for CsPbBr$_3$ QD superlattices, where τ_1, τ_2 are the decay times of the bi-exponential fit and A_1, A_2 the corresponding amplitudes. Assuming that the initial number of coherently coupled QDs is proportional to excitation fluence P_{Exc} with the proportionality constant ζ, the power-dependent effective decay time was fitted with $\tau_{SF}(P_{Exc}) = \tau_{QD}/(\zeta P_{Exc} + 1) + y_0$. Here, a fixed value of $\tau_{QD} = 400$ ps was used, obtained from the time-resolved PL measurements of uncoupled QDs and an additional offset y_0 was inserted to account for effects like the finite time resolution. We obtain good agreement with the expected behaviour ($\tau_{SF} \propto \tau_{QD}/N$ for a value $\zeta_{CsPbBr3} = 0.29 \pm 0.04$ cm^2 μJ^{-1}). In the lower panel of Fig. 5c, we plot τ_D for the CsPbBr$_3$ QD superlattices as a function of the excitation power. In our analysis, the build-up time is composed of the actual delay time due to the SF build-up and a systematic, constant time-offset because the absolute arrival time of the excitation pulse (which has a different wavelength from the emission) at the sample cannot be measured reliably at the required precision from the streak camera data. We observe a decrease in τ_D of about 6 ps when increasing the excitation density by almost 2 orders of magnitude. We have fitted this behaviour for the build-up time τ_D with $\tau_D = y_{offset} + A \ln(\zeta P_{Exc} + 1)/(\zeta P_{Exc} + 1)$ because we assume that $\tau_D \propto \ln(N)/N$ and that the number of excited coupled emitters $N \propto \zeta P_{Exc} + 1$ is proportional to the excitation fluence P_{Exc} with the proportionality constant ζ. Herein, we use a fixed value $\zeta_{CsPbBr3} = 0.29 \pm 0.04$ cm^2 μJ^{-1}, which we obtained from the fit of the effective decay in the upper panel of Fig. 5c, and an amplitude prefactor A. The resulting fit agrees very well with the data. To obtain the absolute time delay, we subtracted the constant offset y_{offset} of the time-delay fit from the time-delay data points. SF occurs when $\sqrt{\tau_{SF}\tau_D} < T_2^*$, where T_2^* is the exciton pure dephasing time, whereas $\sqrt{\tau_{ASE}\tau_D} > T_2^*$ signifies the amplified spontaneous emission (ASE) regime, when τ_{ASE} is the decay time[33]. Considering that the coherence time $T_2 < T_2^*$ extracted from the FWHM of single QDs[11] is of the order of $T_2 = 6.6$ ps, our measurements reveal a fast decay of about 14 ps and a delay time of <1 ps which satisfies the criterion for the appearance of SF.

A similar analysis was performed on CsPbBr$_2$Cl QD superlattices, as shown in Extended Data Fig. 8. In the upper panel of Extended Data Fig. 8c we plot power-dependent effective decay time and fit the data with the same model as described above, using $\tau_{QD} = 250$ ps, and we obtain a value $\zeta_{CsPbBr2Cl} = 0.08 \pm 0.01$ cm^2 μJ^{-1}. The peak intensity of the decay curves as a function of the excitation density is shown in the middle panel of Extended Data Fig. 8c, which increases superlinearly with a power-law dependence with an exponent $\alpha_{CsPbBr2Cl} = 1.3 \pm 0.1$. Also, the build-up time decreases as a function of the excitation density, as displayed in the lower panel of Extended Data Fig. 8c, and fitted the data. Again, we fitted the data using the same formula for τ_D as described above with a fixed value $\zeta_{CsPbBr2Cl} = 0.08 \pm 0.01$ cm^2 μJ^{-1}, and obtained good agreement.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

34. DeVoe, R. G. & Brewer, R. G. Observation of superradiant and subradiant spontaneous emission of two trapped ions. *Phys. Rev. Lett.* **76**, 2049–2052 (1996).
35. Bienaimé, T., Piovella, N. & Kaiser, R. Controlled Dicke subradiance from a large cloud of two-level systems. *Phys. Rev. Lett.* **108**, 123602 (2012).
36. Guerin, W., Rouabah, M. T. & Kaiser, R. Light interacting with atomic ensembles: collective, cooperative and mesoscopic effects. *J. Mod. Opt.* **64**, 895–907 (2017).
37. van Driel, A. F. et al. Statistical analysis of time-resolved emission from ensembles of semiconductor quantum dots: interpretation of exponential decay models. *Phys. Rev. B* **75**, 035329 (2007).
38. Naeem, A. et al. Giant exciton oscillator strength and radiatively limited dephasing in two-dimensional platelets. *Phys. Rev. B* **91**, 121302 (2015).

a

b

Size = 9.45±0.41 nm
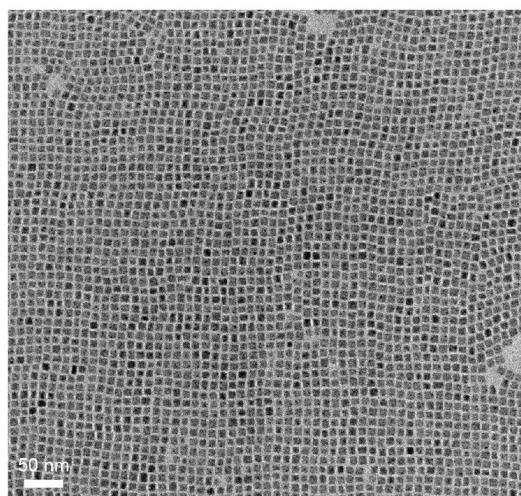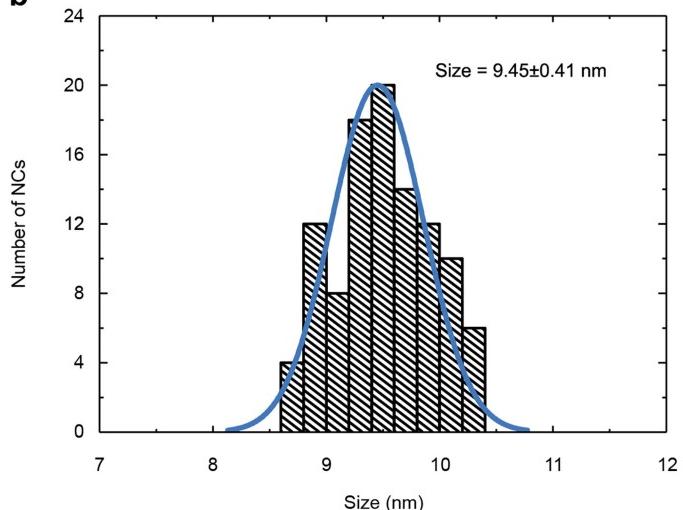
Extended Data Fig. 1 | Quantitative analysis of CsPbBr₃ NC size distribution. a, Low-resolution TEM image of the NC material used to prepare the superlattices. **b**, Histogram of NC sizes (of >100 NCs) as obtained from TEM image analysis. The solid line is a fit with a normal distribution, and the given mean size (9.45 nm) and standard deviation (0.41 nm) are obtained from this fit.

0.1 µm
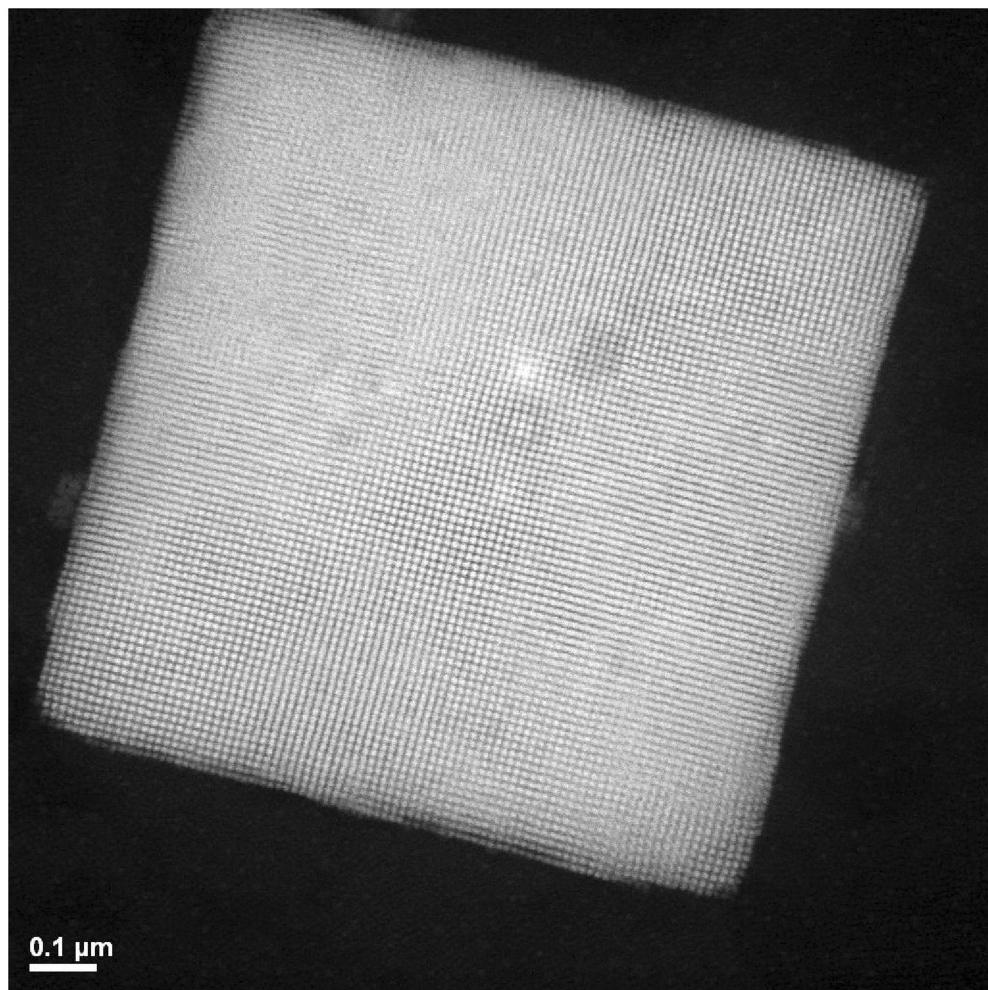
Extended Data Fig. 2 | HAADF-STEM image of a single superlattice of CsPbBr₃ NCs. Individual NCs (bright spots in the image) are well-resolved.
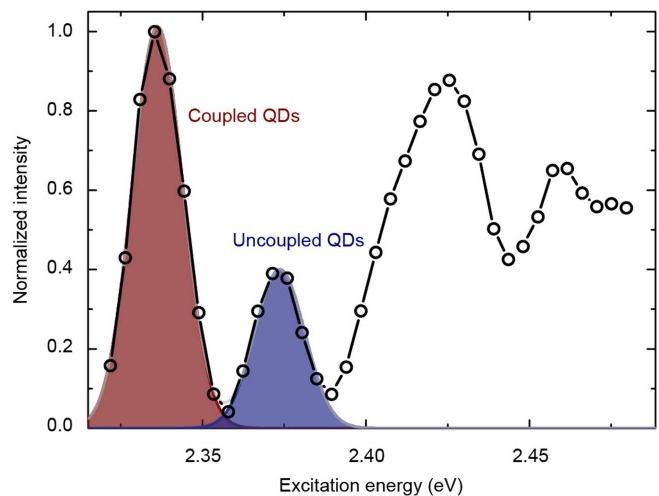
Extended Data Fig. 3 | PL excitation measurement of CsPbBr$_3$ QD superlattices. Using a weak, tunable excitation source, we plot the PL intensity (black circles) obtained at 2.30 eV photon energy as a function of excitation photon energy. The shaded areas are Gaussian peak fits. This PL excitation measurement shows that the coupled QD feature (red) is also present in absorption, in addition to the peak from uncoupled QDs (blue) and energetically higher states.
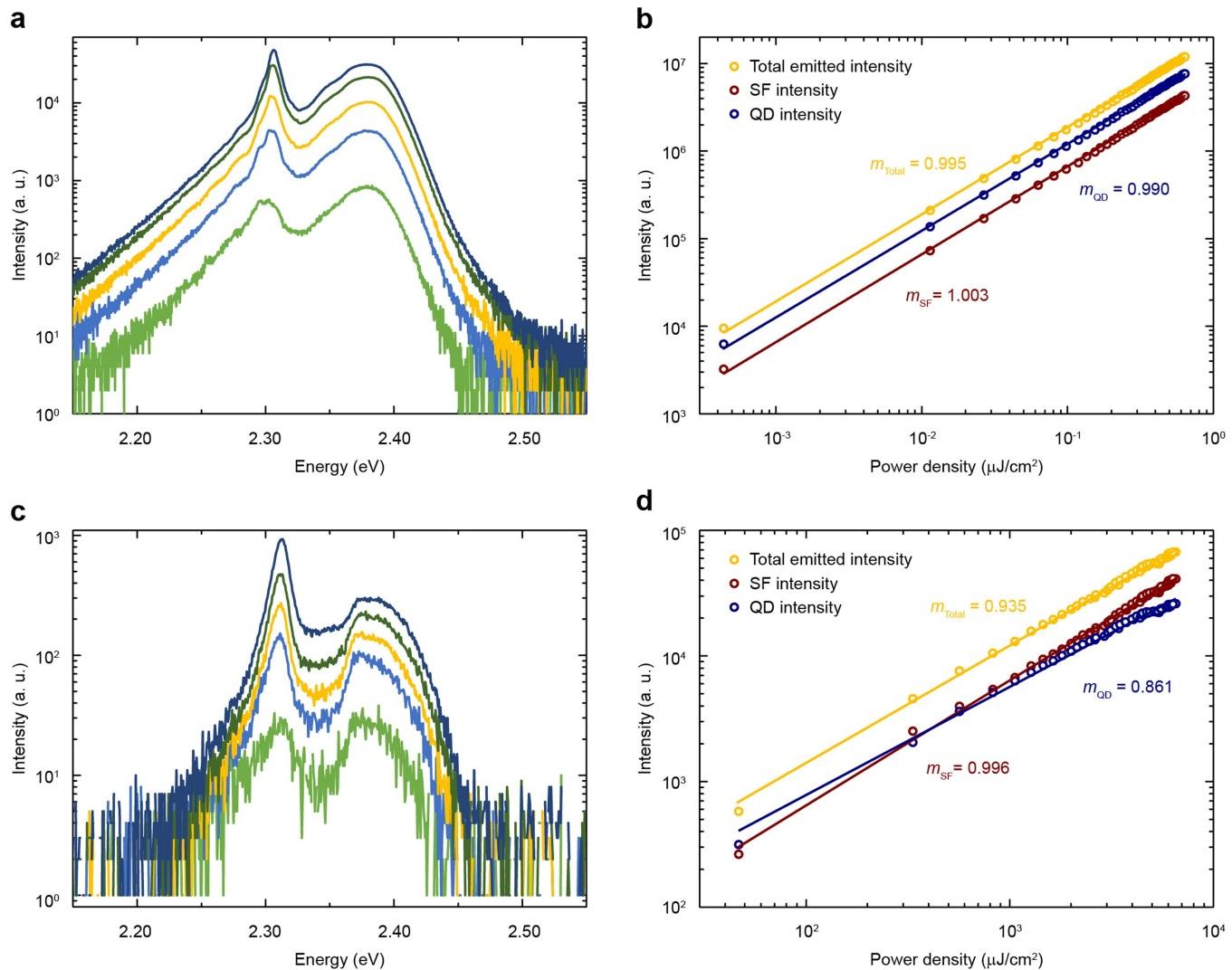
Extended Data Fig. 4 | Power-dependent PL properties of CsPbBr₃ QD superlattices. a, Colour-coded PL emission in the low-power excitation regime, shown for increasing excitation fluence (in nJ cm⁻²) of 10 (light green), 60 (light blue), 150 (yellow) 310 (dark green) and 600 (dark blue). **b,** PL intensity integrated over the spectral emission range of the uncoupled QDs ('QD intensity', blue circles) and coupled QDs ('SF intensity', dark-red circles) in a log–log plot and the total emitted intensity (yellow circles). Fits to the data reveal linear behaviour, as represented by fitted power-law exponents m of approximately 1. **c,** Colour-coded PL emission in the high-power excitation regime, shown for increasing excitation fluence (in μJ cm⁻²) of 330 (light green), 1,270 (light blue), 2,130 (yellow), 3,470 (dark green) and 6,330 (dark blue). **d,** As **b,** but for the high-power excitation regime. Fits to the data reveal a power-law behaviour with a linear increase for the SF emission, a slightly sublinear increase for the uncoupled QDs and a less sublinear increase for the total emitted intensity.

a

b
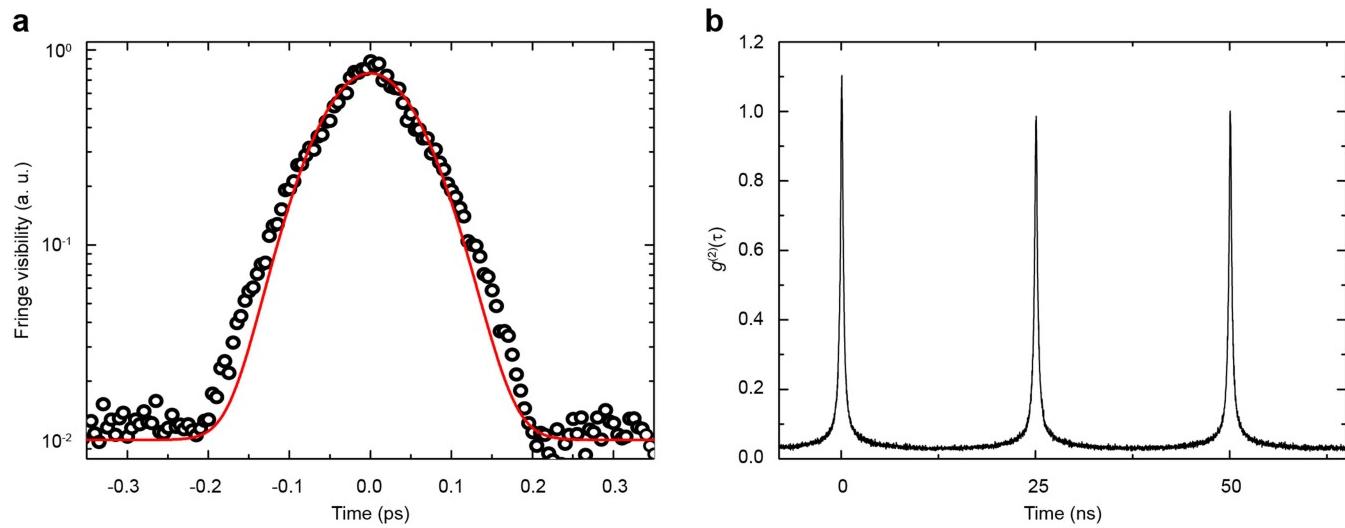
Extended Data Fig. 5 | Gaussian first-order coherence decay and photon bunching in pulsed excitation of CsPbBr$_3$ QD superlattices. a, First-order coherence of the coupled QD emission extracted from the fringe visibility of the interferograms as a function of delay time between the arms of a Michelson interferometer, revealing a mixture of Gaussian (Kubo) and exponential decay (for some of the superlattices). The red solid line is a fit with a mixed Gaussian/exponential model function. **b**, Second-order photon correlation measurement of the coupled QD emission from a single superlattice showing photon bunching at zero delay under pulsed excitation with a 40 MHz repetition rate.
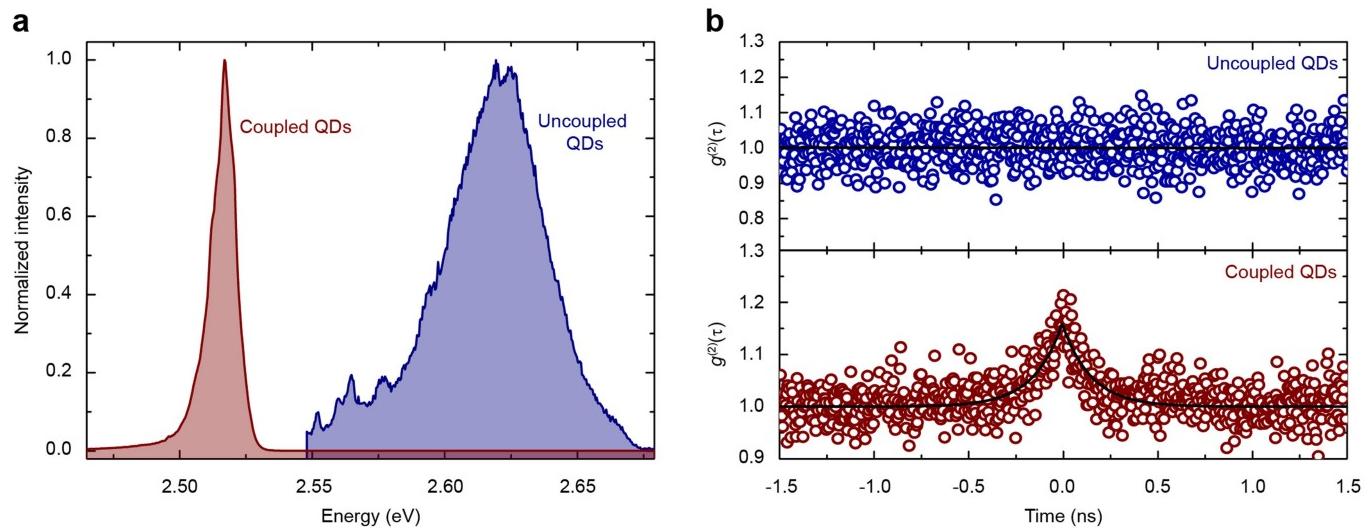
Extended Data Fig. 6 | SF in CsPbBr$_2$Cl QD superlattices. a, Band-pass filtered PL spectra of uncoupled QD emission (blue) and coupled QD emission (red) from CsPbBr$_2$Cl perovskite superlattices. **b**, Second-order photon correlation measurements of uncoupled QD emission (top panel, $g^{(2)}(\tau) = 1$), showing a flat correlation function, and of coupled QD emission (lower panel, $g^{(2)}(0) = 1.15$), showing photon bunching.
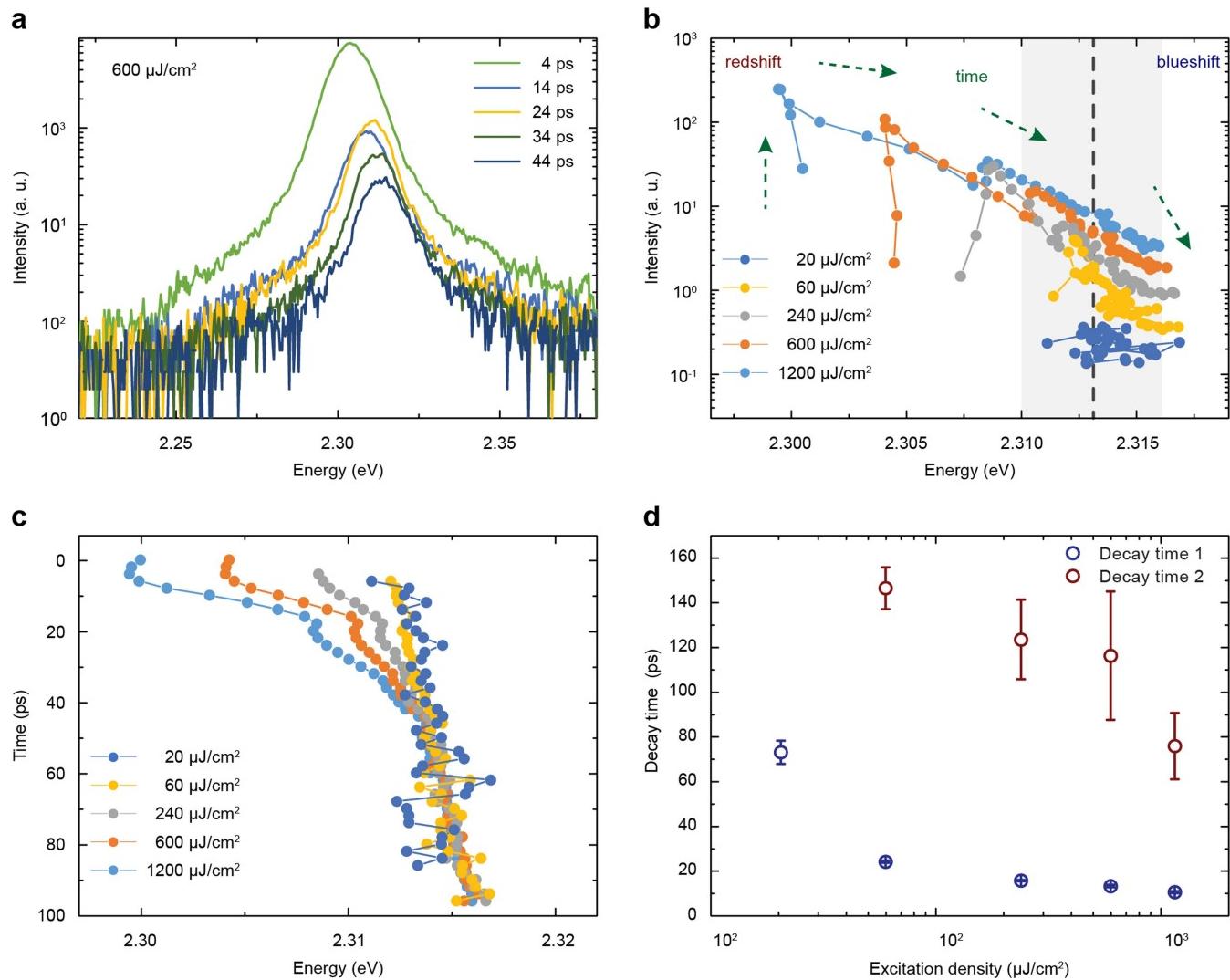
Extended Data Fig. 7 | SF decay and dynamic red-shift of CsPbBr$_3$ QD superlattices. a, PL spectra (integrated over a 2 ps time window) at different time delays on a semi-log scale. **b**, The PL spectra are fitted to a single Gaussian peak function, and the fitted peak amplitude as a function of the emission energy is plotted for various excitation densities (see key). Green arrows indicate the time evolution of the emission peak. The black dashed line denotes the mean energy at the lowest excitation density, and the grey shaded area is the peak's FWHM. **c**, Fitted peak centre energy as a function of time for various excitation densities (see key). **d**, Fast and slow PL decay time components τ_1 and τ_2 of the SF bi-exponential fit model as a function of excitation density on a semi-log plot. The error bars represent the parameters' fit uncertainty.
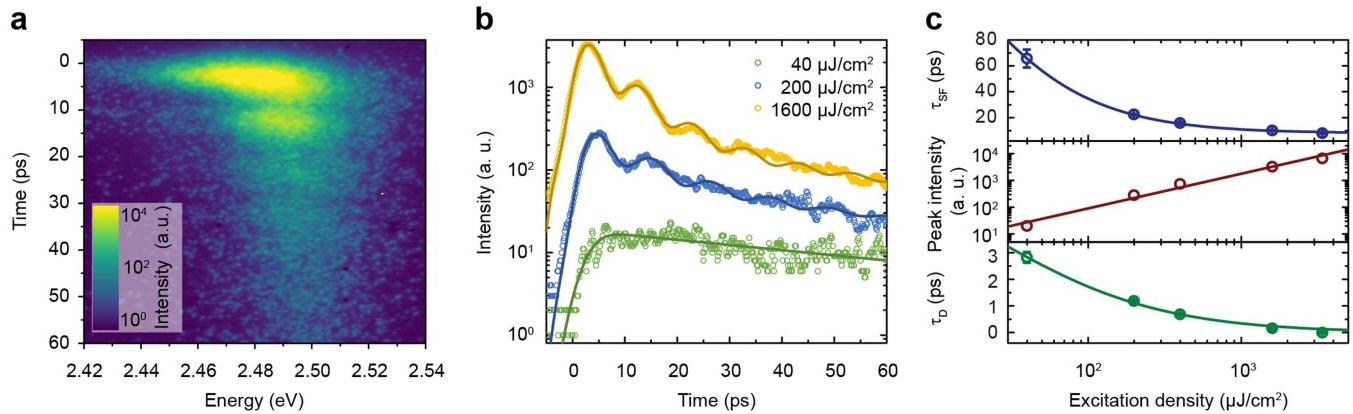
Extended Data Fig. 8 | Burnham–Chiao ringing behaviour in CsPbBr$_2$Cl QD superlattices. a, Streak camera image of SF dynamics obtained with a high excitation density of 1,600 μJ cm^{-2}. **b**, Extracted time-resolved emission intensity traces for three different excitation powers (see key). Solid lines are best-fits to a model that employs a bi-exponential decay function with damped oscillations. **c**, Top, effective SF decay time (blue circles) as a function of the excitation fluence fitted according to the SF model (solid blue line). Middle, dark-red circles represent the peak SF emission intensity that increases superlinearly with excitation power, corresponding to a power-law dependence with an exponent $\alpha = 1.3 \pm 0.1$ (solid dark-red line). Bottom, the extracted delay time τ_D (green circles) decreases at high excitation power due to the increased interaction among the emitters. The green solid line is the best fit according the model described in the Methods section. The error bars represent the parameters' fit uncertainty.
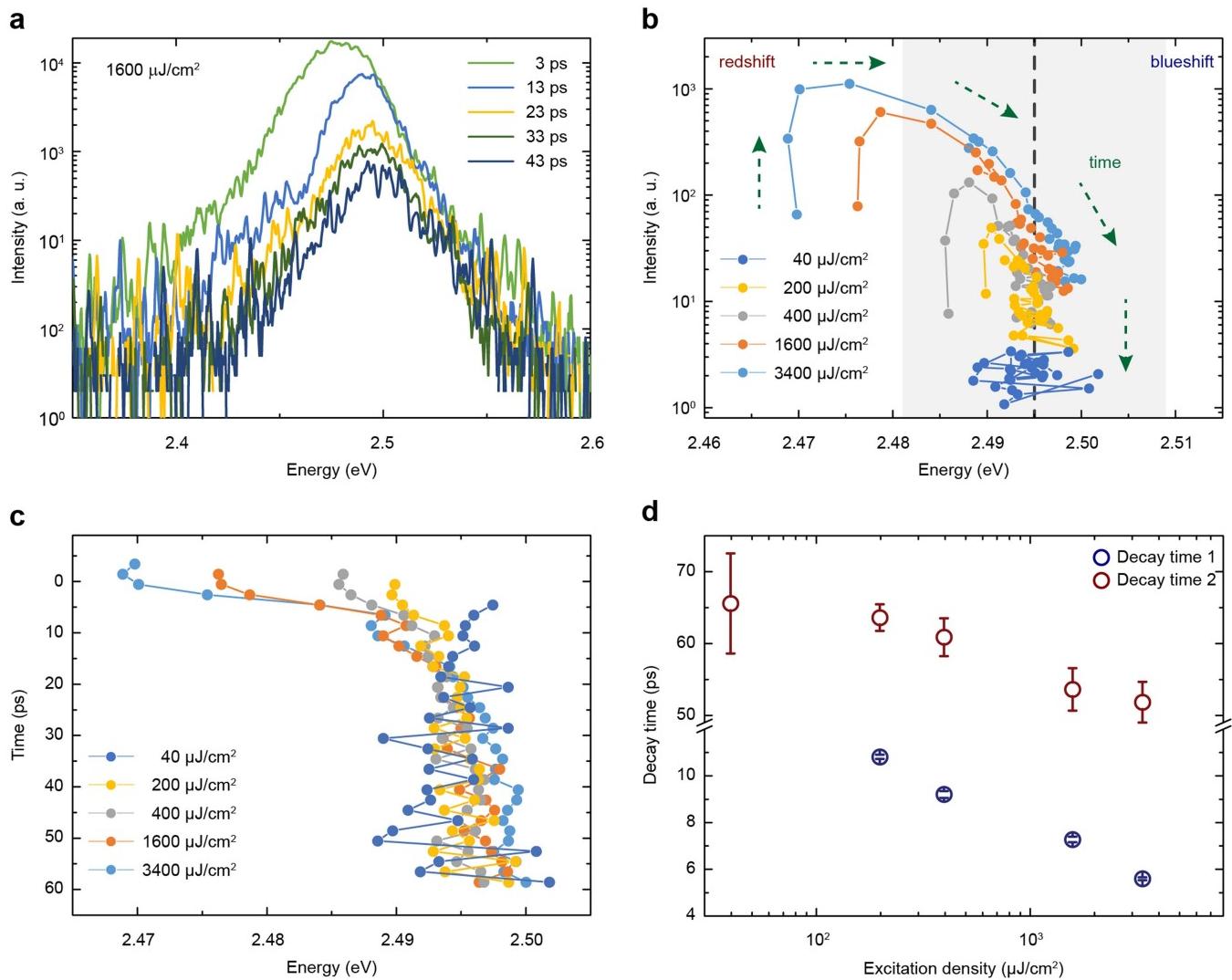
Extended Data Fig. 9 | SF decay and dynamic red-shift of CsPbBr$_2$Cl QD superlattices. a, PL spectra (integrated over a 2 ps time window) at different time delays on a semi-log scale. **b**, The PL spectra are fitted to a single Gaussian peak function, and the fitted peak amplitude as a function of the emission energy is plotted for various excitation densities (see key). Green arrows indicate the time evolution of the emission peak. The black dashed line denotes the mean energy at the lowest excitation density, and the grey shaded area is the peak's FWHM. **c**, Fitted peak centre energy as a function of time for various excitation densities (see key). **d**, Fast and slow PL decay time components τ_1 and τ_2 of the SF bi-exponential fit model as a function of excitation density. The error bars represent the parameters' fit uncertainty.

a

b
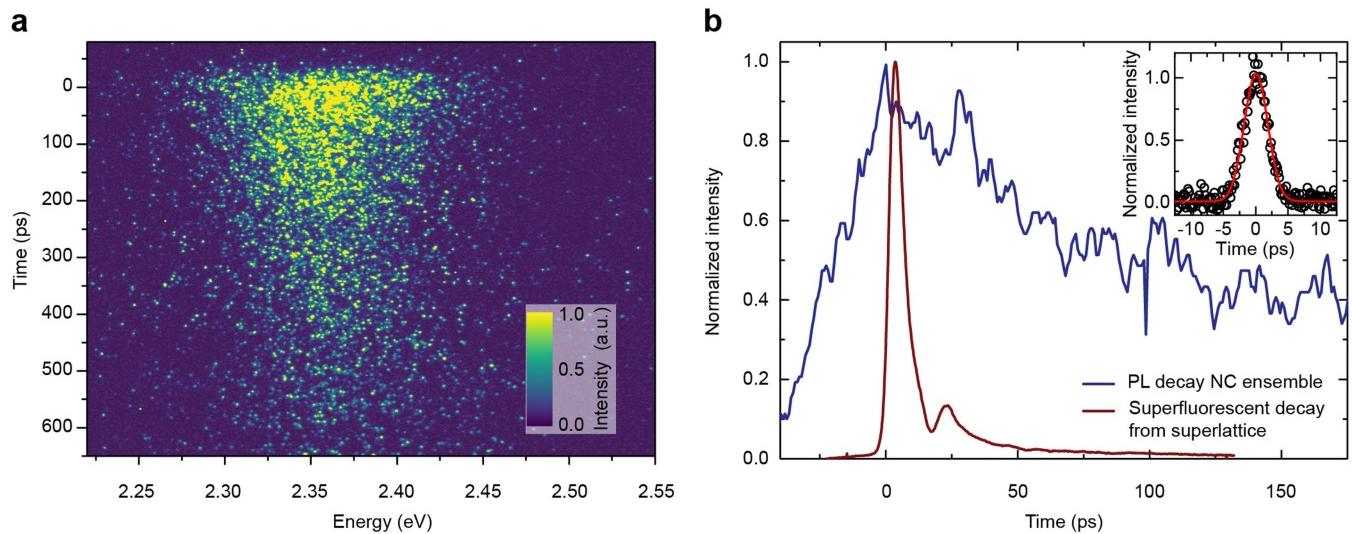
Extended Data Fig. 10 | PL decay from a control sample of diluted CsPbBr₃ QDs. a, Streak camera measurement of a control sample prepared by spin-coating a low concentration of $CsPbBr_3$ QDs dispersed in polystyrene such that no QD coupling is realized. The excitation conditions were comparable to those used in Fig. 5 (an excitation density of 1,600 μJ cm^{-2} was used here). However, the relatively long decay time requires that a longer time range with a lower temporal resolution compared to the measurement in Fig. 5 has to be used on the streak camera. The absence of dynamic red-shift, accelerated decay and ringing proves that the observed SF features cannot be explained by single

QD physics but are due to a multi-particle effect. **b**, Time-resolved PL trace (blue) obtained from these streak data. For direct comparison, the equivalent data from a superlattice sample exhibiting SF (Fig. 5b, 1,200 μJ cm^{-2}) is shown (dark red). Inset, instrument response function of the streak camera. Using the same instrument settings as for the measurement on superlattices, the temporal response of the streak camera using scattered light from a 100-fs laser pulse at 400 nm gives a Gaussian-shaped instrument response with a FWHM of 4.3 ps (black circles); a Gaussian fit is shown as the red solid line.

LETTER

A structural transition in physical networks

Nima Dehmamy[1], Soodabeh Milanlouei[1] & Albert–László Barabási[1,2,3]*

In many physical networks, including neurons in the brain[1,2], three-dimensional integrated circuits[3] and underground hyphal networks[4], the nodes and links are physical objects that cannot intersect or overlap with each other. To take this into account, non-crossing conditions can be imposed to constrain the geometry of networks, which consequently affects how they form, evolve and function. However, these constraints are not included in the theoretical frameworks that are currently used to characterize real networks[5–7]. Most tools for laying out networks are variants of the force-directed layout algorithm[8,9]—which assumes dimensionless nodes and links—and are therefore unable to reveal the geometry of densely packed physical networks. Here we develop a modelling framework that accounts for the physical sizes of nodes and links, allowing us to explore how non-crossing conditions affect the geometry of a network. For small link thicknesses, we observe a weakly interacting regime in which link crossings are avoided via local link rearrangements, without altering the overall geometry of the layout compared to the force-directed layout. Once the link thickness exceeds a threshold, a strongly interacting regime emerges in which multiple geometric quantities, such as the total link length and the link curvature, scale with the link thickness. We show that the crossover between the two regimes is driven by the non-crossing condition, which allows us to derive the transition point analytically and show that networks with large numbers of nodes will ultimately exist in the strongly interacting regime. We also find that networks in the weakly interacting regime display a solid-like response to stress, whereas in the strongly interacting regime they behave in a gel-like fashion. Networks in the weakly interacting regime are amenable to 3D printing and so can be used to visualize network geometry, and the strongly interacting regime provides insights into the scaling of the sizes of densely packed mammalian brains.

To lay out physical networks, the links and nodes must be arranged in such a way to avoid crossing each other, while minimizing the total length of the links, because long links can be costly in systems such as brains. In other words, we must find the shortest path for each link, which may not be a straight path if the straight path is obstructed by other nodes and links—a problem that is equivalent to stretching a rubber band between flexible obstacles (Fig. 1; see Supplementary Information section 3.A for a proof of this equivalence[10]).

To find the shortest path, we propose a model in which the forces that govern the motion of the nodes and links are determined by the gradient of the total potential energy. We define the total potential energy as:

$$V = V_{el} + V_{NL} + V_{NN} + V_{LL}$$
$$= \frac{k}{2} \sum_l \int \left| \frac{dx_l}{ds_l} \right|^2 ds_l + k \sum_{i=1}^N \sum_{l \in \langle i \rangle} X_i \cdot \frac{dx_l}{ds_l} \bigg|_{s_l = s_l^{(end)}} \quad (1)$$
$$+ A_N \sum_{i \neq j} \exp\left(-\frac{|X_i - X_j|^2}{4r_N^2} \right) + A_L \sum_{l \neq m} \int \int \exp\left(-\frac{|x_l - x_{ml}|^2}{4r_L^2} \right) ds_l ds_m$$

where V_{el} is the total elastic potential of all links ($l = 1, \ldots, L$). Each link is modelled as an elastic cylinder with radius r_L, which experiences internal elastic forces and short-range external repulsive forces from other links and nodes; nodes are modelled as spheres. V_{NL} captures the node–link interactions at the endpoints of the links; the non-crossing condition is ensured by a short-range repulsive force in the node–node interaction V_{NN} and in the link–link interaction V_{LL}, which are both modelled as short-range Gaussian potentials with strengths set by A_N and A_L, respectively. In addition, s_l parameterizes the length of link l, with $s_l^{(end)}$ denoting its endpoint; $x_l(s_l, t)$ is the position of a point along the centre of link l at time t; $X_i(t)$ is the position of node i ($i = 1, \ldots, N$); r_N is the range of the node–node repulsive force; k is the elastic constant of the links; and $l \in \langle i \rangle$ indicates that the sum is over all links connected to node i. The potential energy in equation (1) is inspired by models used in self-avoiding polymer chains[11] and manifold dynamics[12]; however, given the constraints induced by the network structure, equation (1) has different terms and describes behaviour that is unique to networks.

With $V_{LL} = 0$ and replacing V_{el} with the elastic energy of a spring, equation (1) reduces to the potential energy of a force-directed layout (FDL) with short-range node repulsion. The lowest-energy solution of equation (1) can involve sharp bending of some links, which we avoid by using a Gay–Berne potential[13], as in polymer physics (Supplementary Information section 4). Finally, we embed the network in a high-viscosity medium, allowing it to relax to a low-energy state without oscillations. Therefore, the node and link positions (X_i and x_l) follow the first-order gradient-descent equations of motion:

$$\lambda_N \frac{dX_i}{dt} = -\frac{\partial V}{\partial X_i}; \quad \lambda_L \frac{dx_l}{dt} = -\frac{\partial V}{\partial x_l} + \frac{d}{ds_l} \frac{\partial V}{\partial (dx_l/ds_l)} \quad (2)$$

where λ_N and λ_L are the friction constants of the nodes and links (Supplementary Information section 3.F). We use an FDL to set the initial positions of the nodes and explore two versions of the model with different constraints: (i) in the elastic-link model (ELI), which corresponds to the limit $\lambda_N \to \infty$, the positions of the nodes are fixed and only the links can reorganize; (ii) in the fully elastic model (FUEL), we assume that $\lambda_N \approx \lambda_L$ and hence the nodes and links are all free to move.

The network defined by equations (1) and (2) has an uneven potential-energy landscape[14] with a very large number of local minima; identifying the globally optimal configuration is NP hard (Supplementary Information section 3.G). We therefore use simulated annealing[15] to approach an energetically favourable local minimum (Supplementary Information section 3.G). The computational complexity of the model is discussed in Supplementary Information section 8.C. In Fig. 1c we show how FUEL finds the optimal three-dimensional configuration of a lattice, helped by the thermal fluctuations from simulated annealing that were added to the links, which allow the layout to tunnel through the finite potential walls and escape local minima.

Because FDLs do not take into account the physical dimensions of the nodes and links, they typically have multiple link and node crossings (Supplementary Information section 2). The number of crossings increases linearly with r_L (Fig. 2a), as predicted analytically by a geometric model (Supplementary Information section 2). To avoid these crossings, we applied ELI and FUEL to several networks with different topologies (random networks and Barabási–Albert[16] scale-free

[1]Network Science Institute, Center for Complex Network Research, Department of Physics, Northeastern University, Boston, MA, USA. [2]Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [3]Department of Network and Data Sciences, Central European University, Budapest, Hungary. *e-mail: alb@neu.edu

a Theory

FDL
Link 2
Link 1

ELI and FUEL

r_L
F_{el} F_{LL}

b Model

FDL

ELI and FUEL

c

Total link length

Simulation time
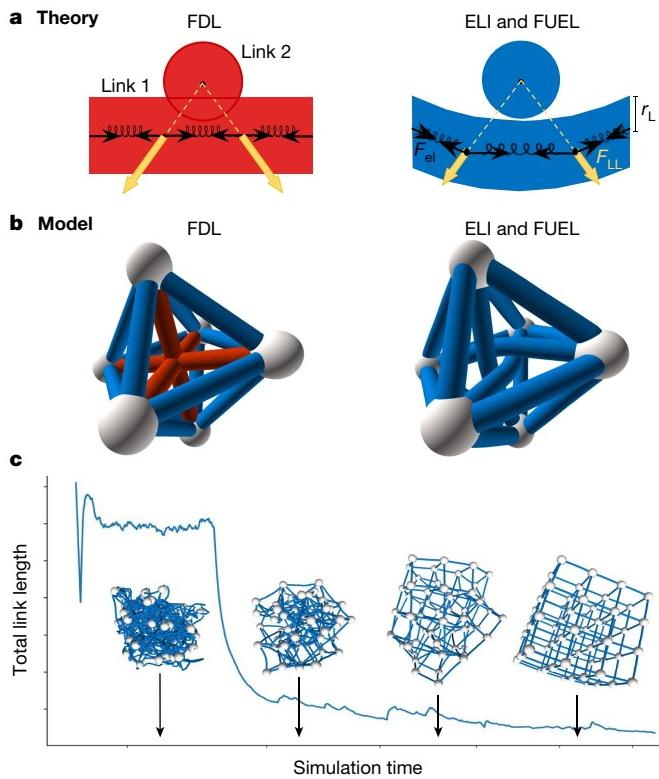
Fig. 1 | Modelling framework to avoid link and node crossings. a, We model each link as a stretched, flexible rubber band, which is represented by many short springs connected to each other, pulled apart by elastic forces F_{el}. In ELI and FUEL, the links exert a repulsive force F_{LL} on each other that falls sharply for radii larger than r_L; for the FDL, $F_{LL} = 0$ and $r_L = 0$. Whereas in the FDL the links may cross each other (left), in ELI and FUEL such crossings are prohibited (right). **b**, A small network with $N = 6$ nodes laid out using the FDL (left), which results in multiple link crossings (red links). Laying the network out using ELI (right) resolves these crossings. **c**, Evolution of the total link length (main plot) and layout of the lattice (insets) during simulated annealing, which determines the final layout of a lattice by minimizing the total link length, starting from a random layout with $r_L \ll r_N$. The thermal noise from the annealing helps links to pass through each other to resolve crossings.

networks), sizes and link densities. We find that the networks undergo a geometric transition as we increase the link thickness (Fig. 2e–h).

For small r_L (the weakly interacting regime), the ELI and FUEL layouts are largely indistinguishable from the initial FDL. At low r_L, the average link length $\langle l \rangle$ is independent of r_L, even as r_L increases by orders of magnitude (Fig. 2b). This is unexpected, given that there is an increase by a factor of ten in the number of potential link crossings in this regime (Fig. 2a). The unchanged $\langle l \rangle$ indicates that ELI and FUEL avoid the increasing number of crossing via only a small amount of local bending of the links. Similar behaviour is seen for the average curvature of the links $\langle C \rangle$. We find that $\langle C \rangle$ changes only modestly from its value at the smallest r_L throughout the weakly interacting regime (Fig. 2c), which indicates that despite the multiple bends in some links that are necessary to avoid crossings the links remain mostly straight. Note that the behaviour of $\langle C \rangle$ in the weakly interacting regime is model-dependent: the movement of nodes in FUEL provides a way of avoiding crossings that requires less curving of the links. Altogether, we find that in the weakly interacting regime local link rearrangements are sufficient to avoid the multiple crossings that are present in the FDL.

Once r_L exceeds a critical value r_L^c (the strongly interacting regime), we observe a marked change in the geometry of the network (Fig. 2f, h). In ELI, with fixed node positions, the links must take long, convoluted routes outside the network to reach their end nodes because they are unable to find sufficient space between the nodes. This change in the link structure is particularly visible in the skeleton of the layout (white

links in Fig. 2f, h). In FUEL, with flexible node positions, the links reach their destination by pushing the nodes away from each other. These changes for ELI and FUEL alter the behaviour of $\langle l \rangle$, which in the strongly interacting regime increases linearly with r_L. The change in link structure also results in relatively large changes in $\langle C \rangle$ at r_L^c; after the transition, $\langle C \rangle$ decreases as $1/r_L$. Despite the different mechanisms that underpin the two models, the scalings of $\langle l \rangle$ and $\langle C \rangle$ in the strongly interacting regime in ELI and FUEL are independent of the network topology. The linear increase in $\langle l \rangle$ and the $1/r_L$ decrease in $\langle C \rangle$ that we observe for both layout models are consistent with isometric scaling, indicating that the layouts in the strongly interacting regime are structurally similar for different r_L to each other if we rescale them by r_L (Supplementary Information section 5.A).

We determine the origin of the transition in the geometry of the networks by estimating the transition point r_L^c. When the links are much thinner than the node repulsion range r_N, the layout is dominated by the repulsive forces between the nodes, which together occupy the volume $V_N = 4\sqrt{2}\, Nr_N^3/3$ (Supplementary Information section 10). When the volume occupied by the links becomes comparable to V_N, the layout must change to accommodate the links. This change induces the transition from the weakly interacting regime to the strongly interacting regime. Taking into account the volume of all nodes and links, we calculate the transition point $\tilde{r}^c = r_L^c/r_N$ to be (Supplementary Information section 10):

$$\tilde{r}^c = \frac{6A^{1/3}}{A^{2/3} + 12B} \tag{3}$$

where

$$A = -12\left(3\langle k^{3/2}\rangle + \sqrt{9\langle k^{3/2}\rangle^2 - 12B^3}\right)$$

$$B = \left(\frac{3}{4}\right)^{1/3} \frac{cL}{N^{2/3}}$$

$$c = \frac{\langle l \rangle}{R}$$

R is the radius of the layout, k is the degree of the nodes and the average $\langle k^{3/2}\rangle$ is taken over the degree distribution of the network. In scale-free and random networks, in the limit $N \to \infty$ we obtain $\tilde{r}^c \approx aL^{-1/2}N^{1/3}$ (Supplementary Information section 10). Given that in many real and model networks $L \approx mN$ for some constant m, we obtain $\tilde{r}^c \propto N^{-1/6}$; therefore, in the limit $N \to \infty$ we find $\tilde{r}^c \to 0$, which implies that the weakly interacting regime is absent in the thermodynamic limit. In other words, in networks with a large number of nodes, the crossings are so numerous that they cannot be ignored. Consequently, the FDL and other currently used layout tools that do not consider link crossings are expected to be inappropriate for large physical networks because the layouts of such networks are dominated by crossings.

Although networks with different N and L transition at different r_L/r_N ratios, if we scale r_L/r_N by $\tilde{r}^c$ the transition occurs near unity for all networks. Using the scaling exponent of the average link length ($\phi(l) = d[\log(\langle l \rangle)]/d[\log(r_L)]$) as the order parameter, the data collapse to a single curve (Fig. 2k), confirming the validity of equation (3). The fact that the transition points of networks with different topologies (scale-free and random networks, lattices and random geometric graphs; Supplementary Information section 11) exhibit similar dependences on r_L suggests that the transition shown in Fig. 2 is independent of the topology and degree distribution of the network.

Analysis of the effects of the size of the network on the scaling of the order parameter (finite-size scaling analysis) indicates that the layout transition occurs over a small, but non-zero range of r_L/r_N, regardless of the network size (Supplementary Information section 11). This result suggests that we are observing a crossover[17,18] from mean-field behaviour ($\phi(l) = 0$) to scaling behaviour ($\phi(l) = 1$). For ELI and FUEL, the weakly interacting regime is well described by an FDL with local
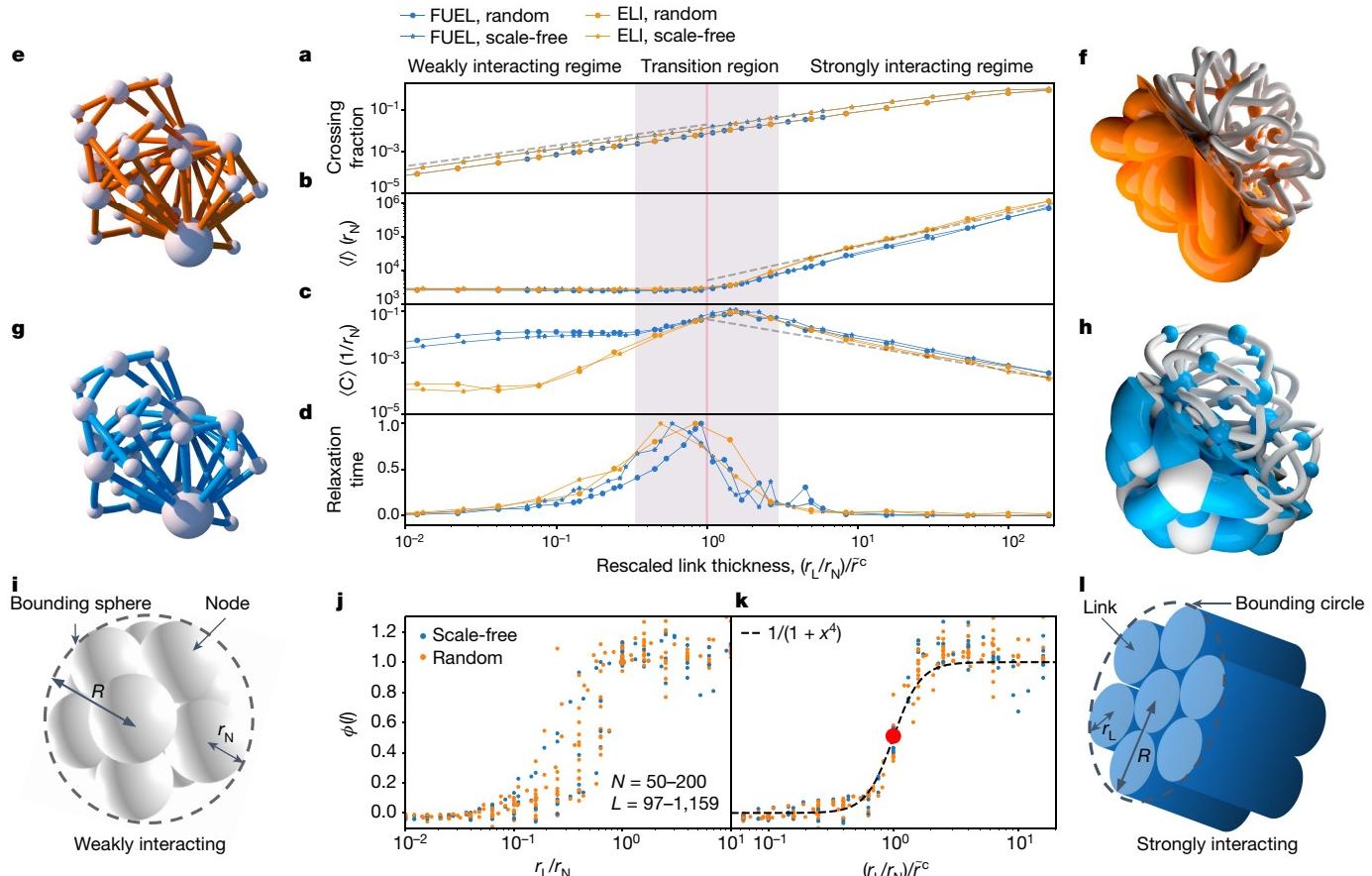
Fig. 2 | Crossover in network layouts. a, The number of link crossings in random (circles) and scale-free (stars) networks using ELI (orange) and FUEL (blue) are calculated assuming that the links are straight (that is, using the FDL), and then normalized by the total number of pairs of links to obtain the crossing fraction shown. The number of link crossings in the FDL grows linearly with r_L (grey dashed line), saturating at very high r_L. A physically realistic layout must resolve this increasing number of crossings. **b**, In all four cases, the average link length $\langle l \rangle$ remains largely constant in the weakly interacting regime, but grows linearly (dashed grey line) in the strongly interacting regime. **c**, The average link curvature $\langle C \rangle$ increases slowly in the weakly interacting regime, with FUEL exhibiting higher average curvature than ELI, then falls linearly (dashed grey line) in the strongly interacting regime. **d**, The relaxation time of the simulated annealing grows substantially near the transition point $\tilde{r}^c = r_L^c/r_N = r_L/r_N$ (vertical pink line). **e–h**, ELI (**e**, **f**; orange) and FUEL (**g**, **h**; blue) layouts for a Barabási–Albert network[16] with $N = 20$ nodes and minimum degree $m = 2$. When $r_L \ll r_N$, the ELI (**e**) and FUEL (**g**) layouts are similar to the FDL (not shown). For larger r_L, links bend to avoid each other: for

ELI (**f**), the links do not fit inside the region containing the nodes and make outward arcs; by contrast, because nodes are free to move for FUEL (**h**), the layout behaves more gently (that is, it contains shorter links, which bend less relative to ELI). In **f** and **h**, the bottom left parts of the image show the full-scale networks and the top right parts show the node and link 'skeletons', with the colours inverted, to help to visualize the geometry. **i**, In the weakly interacting regime ($r_L \ll r_N$), the links are thin and the radius of the entire layout is approximately the radius R of the bounding sphere that surrounds the N nodes of radius r_N. **l**, At larger r_L/r_N, thick links avoid crossing each other and their volume dominates the volume of the whole layout. **j**, The order parameter $\phi(l) = \mathrm{d}[\log(\langle l \rangle)]/\mathrm{d}[\log(r_L)]$ (the scaling exponent of $\langle l \rangle \propto r_L^{\phi(l)}$) versus r_L/r_N for networks with different N (50–200) and L (97–1,159) and different geometries (orange, random; blue, scale-free). **k**, Rescaling the ratio r_L/r_N by $\tilde{r}^c$ (equation (3)) collapses the transition point, shown where $\phi(l) = 1/2$ (red circle). This transition occurs over a small range of r_L/r_N (pink shaded area in **a**–**d**) regardless of the system size, providing evidence of a crossover. The black dashed curve is a smooth fit to the order parameter.

perturbations to resolve possible link crossings. However, this regime disappears in the thermodynamic limit ($N \rightarrow \infty$). In this limit, only the strongly interacting regime is observed, which is dominated by strong link–link interactions and displays universal scaling.

The crossover that we observe also alters the physical properties of the network. For example, the response of a network to external forces is captured by the Cauchy stress tensor[19] $T_{\mu\nu} = \partial_\mu \partial_\nu V$ (Supplementary Information section 6), which depends on the physical and material properties of the nodes and links. In the weakly interacting regime the links are mostly straight; hence, the node terms V_{NN} and V_{NL} dominate the total stress. Because each node is surrounded by a varying number of other nodes, the stress does not spread uniformly in all directions but has shear (off-diagonal) components—a common feature of solids. In the strongly interacting regime, the links fill up the space; hence, the link contributions $V_{el} + V_{LL}$ dominate $T_{\mu\nu}$, resulting in a diagonal total stress tensor (Supplementary Information section 6). In other words, we predict that networks in the strongly interacting regime will display

a fluid or gel-like response to external stress. To test the validity of the solid–gel transition, we compress the networks generated by FUEL in the y direction and measure the tensile forces $\sigma_\mu = T_{\mu\mu}$ (Fig. 3a; Supplementary Information section 6). We again observe a crossover at the value of $\tilde{r}^c$ predicted by equation (3) from a roughly constant stress in the weakly interacting regime to a monotonically increasing stress in the strongly interacting regime (Fig. 3b). Furthermore, as we rotate the network, we find that the stress ratio $\sigma_\parallel/\sigma_\perp$ displays large fluctuations in the weakly interacting regime—behaviour that is often observed in anisotropic solids. The fluctuations vanish at the transition point $\tilde{r}^c$ and the stress ratio settles to the hydrostatic ratio $\sigma_\parallel/\sigma_\perp = 1/\sqrt{2}$ (Fig. 3c)—as expected for gels under pressure.

In summary, the geometry of physical networks is characterized by two distinct regimes: a weakly interacting regime, in which the overlap between the nodes and links is avoided via local link rearrangements, and a strongly interacting regime, the layout of which is shaped by the link–link expulsion. Networks in the weakly interacting regime are
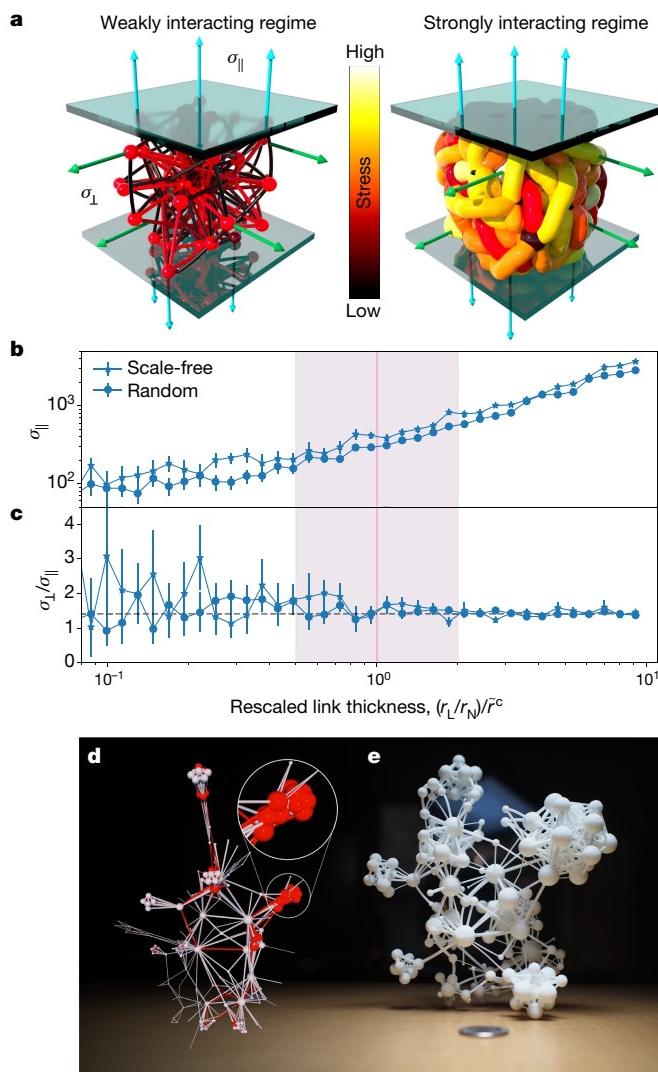
Fig. 3 | Stressed networks and 3D printing. a, The build-up of tensile stress in the nodes and links as a result of compressing the network between two walls. Arrows indicate tensile stress components: cyan, parallel to the direction of compression, $\sigma_\parallel(x)$; green, perpendicular to the direction of compression, $\sigma_\perp(x)$. The networks are coloured on the basis of the total amount of stress. In the weakly interacting regime (left), the stress is concentrated in the nodes; in the strongly interacting regime (right), almost all of the stress is in the links. **b**, The parallel stress component $\sigma_\parallel$ of scale-free (squares) and random (circles) networks as a function of scaled link thickness $(r_L/r_N)/\bar{r}^c$. Because the definition of x, y and z is frame-dependent, we average the forces over 50 random network orientations. **c**, The ratio of parallel and transverse tensile stress components $\sigma_\perp/\sigma_\parallel$. Error bars in **b** and **c** correspond to one standard deviation around the mean, calculated over the 50 random orientations. In the weakly interacting regime, the ratio depends on the orientation of the layout (as can be seen from the large error bars from averaging over the orientations), which indicates solid-like behaviour. In the strongly interacting regime, the fluctuations in $\sigma_\perp/\sigma_\parallel$ decay, yielding a constant ratio. **d**, **e**, Visualization of networks. As an example, we consider a network with $N = 184$ and $L = 716$ that represents ingredients that share flavour compounds[21]. A three-dimensional rendering of the FDL (**d**) results in multiple crossing (red). The inset in **d** highlights a densely connected region (corresponding to dairy products) with a lot of overlap; consequently, it is difficult to discern the underlying network. By contrast, when laying out the flavour network using FUEL (**e**; printed using a commercial 3D printer), the crossings disappear, unveiling the inner structure of the network.

solid-like, whereas those in the strongly interacting regime behave like gels. The transition that we observe between the two regimes is unique to three dimensions: because links are effectively one-dimensional

objects, the non-crossing condition results in knot-like constraints in three dimensions, which prevent the links from passing through each other. In four dimensions or more, knots of one-dimensional objects can be untied[20], so the non-crossing conditions will not constrain the geometry. Therefore, three is the lowest number of dimensions in which links can avoid each other by bending and the highest in which they cannot pass by each other without breaking or tunnelling.

Both regimes have applications. In contrast to the physical networks considered thus far in which the nodes and links have physical sizes, many networks, such as disease–gene interactions, are more abstract, with no real three-dimensional manifestation. In such cases, the layout of the network is not limited by the physical constraints of the system, but can be chosen in such a way to best visualize the underlying network structure. Thus, the weakly interacting regime is appropriate for network visualizations because it clearly separates nodes and links and is amenable to 3D printing, which provides a way of interacting with the network and exploring its inner structure directly. As an example, we consider a network with 184 nodes and 716 links that represents ingredients that share flavour compounds[21]. For networks such as this with high link densities, two-dimensional visualizations suffer from visual cluttering, making only a fraction of the links visible[21]. A three-dimensional layout may provide more clarity, but the FDL still exhibits node and link overlap (Fig. 3d), obstructing the details of the geometry of the network. By contrast, when applying FUEL and choosing r_L to be sufficiently small that the layout is in the weakly interacting regime, we obtain a geometry that reveals the underlying structure of the network and is amenable to 3D printing (Fig. 3e). Given that for large N link crossings in the FDL are inevitable, the method introduced here to resolve crossings will be essential as we aim to visualize large networks. Although the weakly interacting regime vanishes in the thermodynamic limit ($N \to \infty$), for a large but finite network with a fixed number of nodes we will always be able to choose r_L and r_N so that we stay in the weakly interacting regime.

The strongly interacting regime is directly relevant to the brain—a three-dimensional physical network in which the close-packing of the axons is critical to their ability to form synapses[22,23]. A scaling law of $V_w \propto A_w^{1.5}$ between the volume V_w and surface area A_w of the white matter in rodent brains has been observed previously[24]. This law implies that in these networks the average neuron length scales with the axon thickness as $\langle l \rangle = V_w / A_w \propto r_L$, as predicted for the strongly interacting regime (Fig. 2b). If we describe anatomical regions as nodes and axon bundles connecting the anatomical regions as links, then the thickness of the axon bundles r_L is comparable to the size of the anatomical regions. This result supports the prediction of the empirical scaling that these brain networks are in the strongly interacting regime. Thus, equations (1) and (2) provide an appropriate modelling framework to capture the geometry of dense neuronal networks, generating a layout that minimizes the total link length[25,26] while respecting the non-crossing conditions that axons must obey[1].

Data availability

All data used in the figures were generated using the simulation code available at https://github.com/nimadehmamy/3D-ELI-FUEL. The data that support the findings of this study are available from the corresponding author on reasonable request.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0726-6.

1. Kasthuri, N. et al. Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–661 (2015).
2. Oh, S. W. et al. A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).

3. Wong, S. et al. Monolithic 3D integrated circuits. In *2007 International Symposium onVLSI Technology, Systems and Applications* 1–4 (IEEE, 2007).
4. Friese, C. F. & Allen, M. F. The spread of Va mycorrhizal fungal hyphae in the soil: inoculum types and external hyphal architecture. *Mycologia* **83**, 409–418 (1991).
5. Barrat, A., Barthelemy, M. & Vespignani, A. *Dynamical Processes on Complex Networks* (Cambridge Univ. Press, Cambridge, 2008).
6. Barabási, A.-L. *Network Science* (Cambridge Univ. Press, Cambridge, 2016).
7. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
8. Kamada, T. & Kawai, S. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.* **31**, 7–15 (1989).
9. Fruchterman, T. M. & Reingold, E. M. Graph drawing by force-directed placement. *Softw. Pract. Exper.* **21**, 1129–1164 (1991).
10. Dubrovin, B., Fomenko, A. & Novikov, S. *Modern Geometry—Methods and Applications. Part II: The Geometry and Topology of Manifolds [transl. by R. G. Burns]* 371–379 (Springer, New York, 1984).
11. des Cloizeaux, J. Lagrangian theory for a self-avoiding random chain. *Phys. Rev. A* **10**, 1665–1669 (1974).
12. Mézard, M. & Parisi, G. Replica field theory for random manifolds. *J. Phys. I* **1**, 809–836 (1991).
13. Gay, J. & Berne, B. Modification of the overlap potential to mimic a linear site–site potential. *J. Chem. Phys.* **74**, 3316–3319 (1981).
14. Bouchaud, J.-P., Cugliandolo, L. F., Kurchan, J. & Mezard, M. in *Spin Glasses and Random Fields* (ed. Young, A. P.) 161–223 (World Scientific, Singapore, 1998).
15. Kirkpatrick, S., Gelatt, C. D., Jr & Vecchi, M. P. in *Spin Glass Theory and Beyond* (eds Mézard, M., Parisi, G. & Virasoro, M. A.) 339–348 (World Scientific, Singapore, 1987).
16. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
17. Chaikin, P. M. & Lubensky, T. C. *Principles of Condensed Matter Physics* Ch. 5.8 (Cambridge Univ. Press, Cambridge, 2000).
18. Cardy, J. *Scaling and Renormalization in Statistical Physics* Vol. 5, 67–71 (Cambridge Univ. Press, Cambridge, 1996).
19. Irgens, F. *Continuum Mechanics* 60–73 (Springer, Berlin, 2008).
20. Zeeman, E. C. Unknotting combinatorial balls. *Ann. Math.* **78**, 501–526 (1963).
21. Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P. & Barabási, A.-L. Flavor network and the principles of food pairing. *Sci. Rep.* **1**, 196 (2011).
22. Stepanyants, A., Hof, P. R. & Chklovskii, D. B. Geometry and structural plasticity of synaptic connectivity. *Neuron* **34**, 275–288 (2002).
23. Rivera-Alba, M. et al. Wiring economy and volume exclusion determine neuronal placement in the *Drosophila* brain. *Curr. Biol.* **21**, 2000–2005 (2011).
24. Ventura-Antunes, L., Mota, B. & Herculano-Houzel, S. Different scaling of white matter volume, cortical connectivity, and gyrification across rodent and primate brains. *Front. Neuroanat.* **7**, 3 (2013).
25. Bullmore, E. & Sporns, O. The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349 (2012).
26. Sporns, O., Chialvo, D. R., Kaiser, M. & Hilgetag, C. C. Organization, development and function of complex brain networks. *Trends Cogn. Sci.* **8**, 418–425 (2004).

Author contributions N.D. developed, ran and analysed the simulations, performed the mathematical modelling and derivations, and contributed to writing the manuscript. S.M. contributed to programming and running the simulations, generating figures, editing and 3D printing. A.-L.B. contributed to the conceptual design of the study and was the lead writer of the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0726-6.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.-L.B.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

LETTER

Abrupt ice–age shifts in southern westerly winds and Antarctic climate forced from the north

Christo Buizert[1]*, Michael Sigl[2,14], Mirko Severi[3], Bradley R. Markle[4], Justin J. Wettstein[1,5], Joseph R. McConnell[6], Joel B. Pedro[7,8], Harald Sodemann[5], Kumiko Goto-Azuma[9], Kenji Kawamura[9], Shuji Fujita[9], Hideaki Motoyama[9], Motohiro Hirabayashi[9], Ryu Uemura[10], Barbara Stenni[11], Frédéric Parrenin[12], Feng He[1,13], T. J. Fudge[4] & Eric J. Steig[4]

The mid-latitude westerly winds of the Southern Hemisphere play a central role in the global climate system via Southern Ocean upwelling[1], carbon exchange with the deep ocean[2], Agulhas leakage (transport of Indian Ocean waters into the Atlantic)[3] and possibly Antarctic ice-sheet stability[4]. Meridional shifts of the Southern Hemisphere westerly winds have been hypothesized to occur[5,6] in parallel with the well-documented shifts of the intertropical convergence zone[7] in response to Dansgaard–Oeschger (DO) events— abrupt North Atlantic climate change events of the last ice age. Shifting moisture pathways to West Antarctica[8] are consistent with this view but may represent a Pacific teleconnection pattern forced from the tropics[9]. The full response of the Southern Hemisphere atmospheric circulation to the DO cycle and its impact on Antarctic temperature remain unclear[10]. Here we use five ice cores synchronized via volcanic markers to show that the Antarctic temperature response to the DO cycle can be understood as the superposition of two modes: a spatially homogeneous oceanic 'bipolar seesaw' mode that lags behind Northern Hemisphere climate by about 200 years, and a spatially heterogeneous atmospheric mode that is synchronous with abrupt events in the Northern Hemisphere. Temperature anomalies of the atmospheric mode are similar to those associated with present-day Southern Annular Mode variability, rather than the Pacific–South American pattern. Moreover, deuterium-excess records suggest a zonally coherent migration of the Southern Hemisphere westerly winds over all ocean basins in phase with Northern Hemisphere climate. Our work provides a simple conceptual framework for understanding circum-Antarctic temperature variations forced by abrupt Northern Hemisphere climate change. We provide observational evidence of abrupt shifts in the Southern Hemisphere westerly winds, which have previously documented[1–3] ramifications for global ocean circulation and atmospheric carbon dioxide. These coupled changes highlight the necessity of a global, rather than a purely North Atlantic, perspective on the DO cycle.
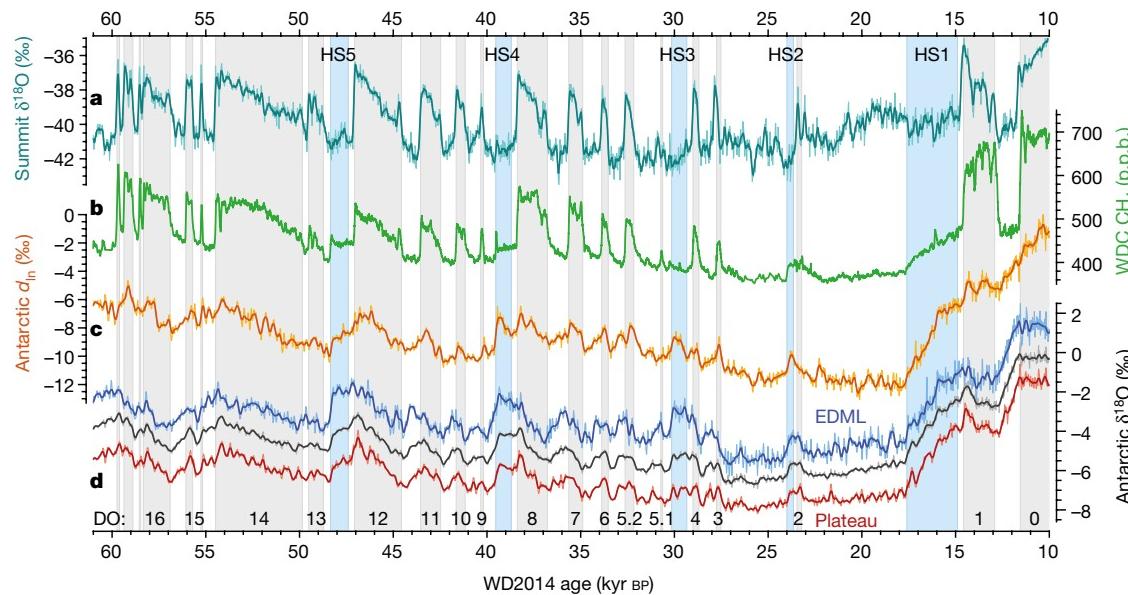
Fig. 1 | Records of abrupt glacial climate variability. a, Ice-core $\delta^{18}O$ (average of GISP2 and GRIP[28]) at the Greenland summit. **b**, WDC methane[29]. p.p.b., parts per 10^9. **c**, Antarctic five-core (EDC, EDML, TAL and DF) average d_{ln} anomaly. **d**, Antarctic $\delta^{18}O$ anomaly at EDML (blue), the Antarctic Plateau (average of DF and EDC; red) and five-core average (black). All records are synchronized to WD2014 chronology; Antarctic data are shown as anomalies relative to the present, with the EMDL and Antarctic Plateau lines offset by +1.4‰ and −1.3‰, respectively, for clarity. DO interstadial periods are marked in grey and numbered, and Heinrich stadials (HS1–HS5) are marked in blue. Isotope ratios are on the VSMOW (Vienna Standard Mean Ocean Water) scale. Thin lines show records at the original resolution (ranging from about 5 to about 50 yr), and thick lines are a moving average (with a 300-yr and 150-yr window for Antarctic and Greenland data, respectively).

[1]College of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA. [2]Laboratory of Environmental Chemistry, Paul Scherrer Institute, Villigen, Switzerland. [3]Department of Chemistry 'Ugo Schiff', University of Florence, Florence, Italy. [4]Department of Earth and Space Science, University of Washington, Seattle, WA, USA. [5]Geophysical Institute and Bjerknes Centre for Climate Research, University of Bergen, Bergen, Norway. [6]Desert Research Institute, Nevada System of Higher Education, Reno, NV, USA. [7]Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark. [8]Antarctic Climate & Ecosystems Cooperative Research Centre, University of Tasmania, Hobart, Tasmania, Australia. [9]National Institute for Polar Research, Tachikawa, Tokyo, Japan. [10]Department of Chemistry, Biology and Marine Science, University of the Ryukyus, Okinawa, Japan. [11]Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, Italy. [12]Université Grenoble Alpes, CNRS, IRD, IGE, Grenoble, France. [13]Center for Climatic Research, Nelson Institute for Environmental Studies, University of Wisconsin-Madison, Madison, WI, USA. [14]Present address: Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland. *e-mail: buizertc@oregonstate.edu
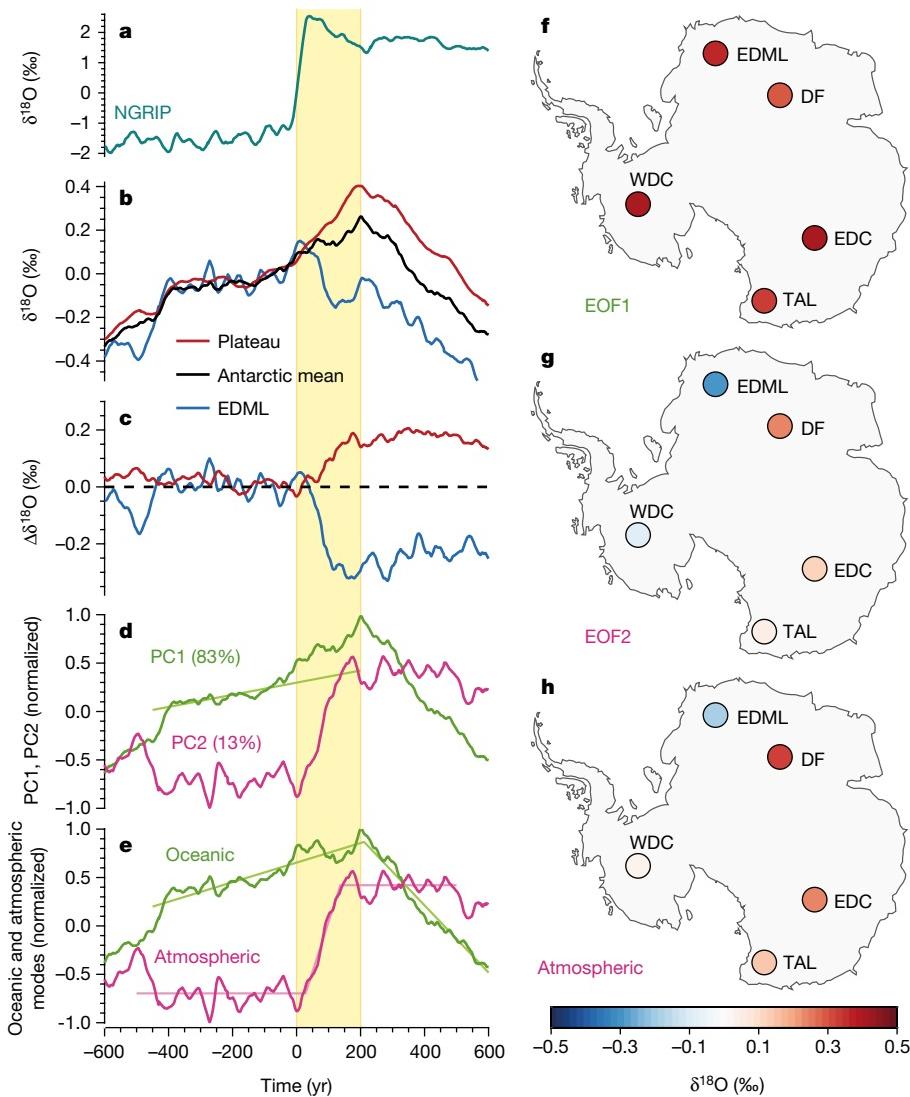
Fig. 2 | Antarctic climate response to DO warming. a, Stack of North Greenland Ice Core Project (NGRIP) $\delta^{18}O$ records. **b**, Stack of Antarctic $\delta^{18}O$ records at the indicated locations, with 'Plateau' representing the average of DF and EDC. **c**, As in **b**, but with the five-core mean subtracted. **d**, First two principal components of the Antarctic $\delta^{18}O$ stacks (1,500-yr window), with the percentage of variance provided (PC2 is offset by −0.15 for clarity). PC1 is strongly correlated ($r = 0.998$) to the Antarctic mean. A linear fit to PC1 (interval from $t = -400$ yr to $t = 0$) is shown to highlight the response around $t = 0$. **e**, Proposed oceanic and atmospheric modes,

obtained using rotated principal component vectors (Methods). Fits from change-point analysis (Methods) are also shown. The oceanic mode responds at $t = 211 \pm 42$ yr and the atmospheric mode at $t = 28 \pm 40$ yr (1σ bounds; Extended Data Table 1). Curves in **d** and **e** are normalized to have $2\sigma = 1$. **f**, **g**, Empirical orthogonal functions EOF1 and EOF2 associated with PC1 and PC2 in **d**, scaled to $\delta^{18}O$ units (‰). **h**, Spatial pattern associated with the atmospheric mode shown in **e**, scaled to $\delta^{18}O$ units (‰). Isotope ratios are on the VSMOW scale.

During the glacial DO cycle, abrupt variations in northward heat transport by the Atlantic meridional overturning circulation (AMOC) affect Greenland and Antarctic temperatures oppositely (Fig. 1), via an oceanic teleconnection called the bipolar seesaw[6,11]. Antarctica warms during Greenland cold phases (stadials) and cools during Greenland warm phases (interstadials), with the gradual nature of Antarctic climate change reflecting buffering by a large heat reservoir[11]—probably the global ocean interior[6]. The DO cycle also affects atmospheric circulation; the intertropical convergence zone shifts southwards during stadials and northwards during interstadials[7]. General circulation model (GCM) simulations suggest parallel shifts of the Southern Hemisphere westerly winds (westerlies)[5,6,12], but the available observational evidence (a deuterium-excess record from West Antarctica[8]) cannot distinguish between such shifts and Pacific-only teleconnections[9]. Furthermore, the impact of the atmospheric circulation changes on Antarctic climate remains unknown, and models are inconclusive on this question[10,13].

We use water stable isotope ratios, a proxy for site temperature[14], from five Antarctic ice cores: WAIS (West Antarctic Ice Sheet) Divide (WDC),

EPICA (European Project for Ice Coring in Antarctica) Dronning Maud Land (EDML), EPICA Dome C (EDC), Dome Fuji (DF) and Talos Dome (TAL). WDC is synchronized to Greenland ice cores at high precision via atmospheric methane (Fig. 1a, b)[15]; here we synchronize WDC to the other cores in the interval 10–57 kyr before present (BP) using volcanic markers (Methods; Extended Data Fig. 1), greatly improving our ability to study the timing of regional Antarctic climate variations relative to Greenland. We investigate the Antarctic response to DO events using a stacking technique[15] in which 19 individual events are aligned at the midpoint of their abrupt methane transition in WDC and averaged to obtain the shared climatic signal (Methods).

Antarctica cools in response to DO warming (Fig. 2a, b), consistent with the bipolar seesaw theory[6,11]. In the Antarctic mean $\delta^{18}O$ stack (where $\delta^{18}O$ represents the $^{18}O/^{16}O$ composition), the cooling onset occurs about two centuries after the abrupt Northern Hemisphere event, providing validation of earlier results from West Antarctica[15]. There is a spatial pattern to the Antarctic response, however. A step-like divergence from the mean signal is seen around $t \approx 0$ yr (that is,
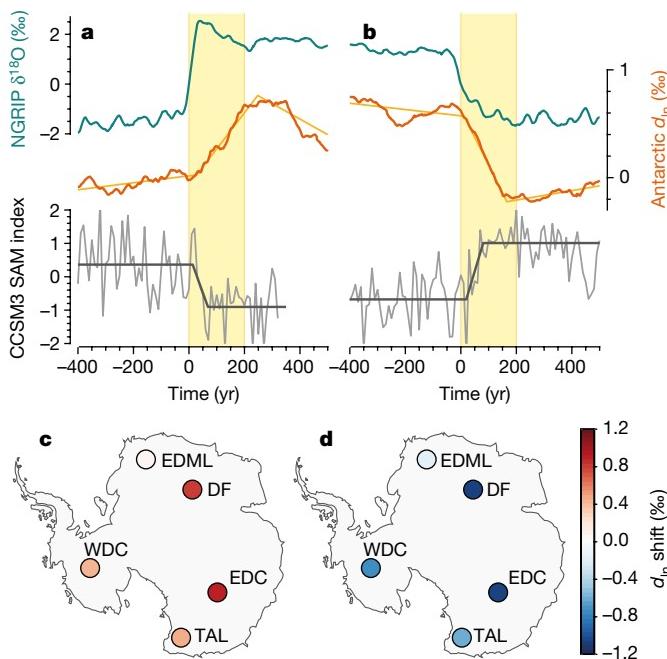
Fig. 3 | Deuterium excess and the Southern Hemisphere westerlies. **a**, DO warming. Shown are data from the NGRIP $\delta^{18}O$ stack (turquoise); five-core average Antarctic d_{ln} stack (orange, with fit from change-point analysis; see Extended Data Table 1); the SAM index (here the leading principal component of sea-level pressure variability south of 20° S) following a freshwater-forced AMOC perturbation in CCSM3 (Community Climate System Model version 3) simulations (grey, with fit from change-point analysis; see Extended Data Table 1). **b**, As in **a**, but for DO cooling. **c**, Magnitude of Antarctic d_{ln} response to DO warming. The weak d_{ln} trend before and after the abrupt jump probably reflects the SST of Southern Hemisphere vapour source waters following the thermal bipolar seesaw[8,11]. **d**, As in **c**, but for DO cooling. Isotope ratios are on the VSMOW scale.

synchronous with Northern Hemisphere climate), with the interior East Antarctic Plateau sites (DF and EDC) warming and EDML cooling (Fig. 2c). This instantaneous warming over the plateau is particularly pronounced at DO events 1, 8, 12 and 14 (Fig. 1d, red curve).

Using principal component analysis (PCA; see Methods), we find that two modes of variability explain more than 96% of signal variance in the five stacked records (Fig. 2d). The first principal component (PC1; 83% of variance explained) has the triangular shape of the Antarctic isotope maximum events—the classic thermal bipolar seesaw signal[11]—with a spatially homogeneous expression (Fig. 2f). The two-century lag behind Greenland warming identifies PC1 as an ocean-propagated response[15].

The second principal component (PC2; 13% of variance explained) is a step-like function with a heterogeneous spatial pattern (Fig. 2g). This mode is very different from the bipolar seesaw. The PC2 response is synchronous with Northern Hemisphere warming within precision (28 ± 40 year lag, 1σ bounds; see Methods); this timing, as well as additional evidence presented below, suggests that this mode represents an atmospheric teleconnection. The PCA does not necessarily separate physical processes. We assume two underlying teleconnections: oceanic (two-century lag) and atmospheric (synchronous). Some amount of each process is included in PC1, as evident by some immediate warming around $t = 0$. We perform a rotation of the PCA vectors (Methods) to isolate the 'purely' oceanic and atmospheric responses (Fig. 2e). The associated estimate of the atmospherically forced temperature anomaly (Fig. 2h) is cooling at EDML, warming at DF, EDC and TAL, and a negligible response at WDC; this pattern is robustly reproduced using different methods (Extended Data Fig. 6). The magnitude of the Antarctic atmospheric response is roughly proportional to the $\delta^{18}O$ perturbation in Greenland ice cores (Extended Data Fig. 4).

The Antarctic response to DO cooling is qualitatively similar to the DO warming case. The ocean seesaw warming response is delayed by 226 ± 44 yr and the spatial pattern of the empirical orthogonal function EOF2 has the opposite sign—that is, additional warming at EDML and cooling on the interior of the East Antarctic Plateau (Extended Data Fig. 7). The atmospheric signal over Antarctica is much weaker for the DO cooling case, with PC2 explaining only 9% of variance. This difference is likely due to the fact that DO warmings are more abrupt and of larger magnitude than DO coolings.

To better understand the atmospheric mode, we turn to deuterium excess (d), a proxy for vapour source conditions[16] commonly used to identify changes in atmospheric circulation and vapour transport pathways[8,17,18]. In isotope-enabled GCM simulations, Antarctic d is anti-correlated with the Southern Annular Mode (SAM) index[8,19]. This anti-correlation can be understood conceptually: when the Southern Hemisphere westerlies are displaced equatorwards (negative SAM phase), Antarctic moisture will originate from further north, where sea surface temperature (SST) is higher and relative humidity is lower (Extended Data Fig. 8b), both of which act to make d more positive[16]. We use the logarithmic definition of deuterium excess (d_{ln}), which preserves isotopic moisture source information better than the linear definition[8,20].

The Antarctic mean d_{ln} response (Fig. 3a, b) lags behind Northern Hemisphere climate by 8 ± 48 yr for DO warming and 9 ± 42 yr for DO cooling, consistent with previous results for WDC[8]. The observed d_{ln} response is consistent with a shift in the meridional position of the Southern Hemisphere westerly winds and vapour origin, such that they move equatorwards in response to Northern Hemisphere warming and polewards in response to cooling. The timing suggests propagation to high latitudes of the Southern Hemisphere via an atmospheric tele-connection. The d_{ln} response is largest for the interior Plateau sites (DF and EDC), possibly because their vapour source areas are more distant from confounding local effects such as the sea–ice edge[21]. The response is weak or absent at EDML, possibly because the variability of Southern Hemisphere westerlies is relatively weak in the Atlantic sector (Extended Data Fig. 9), or because of regional effects, such as wind-driven changes to the sea ice, gyre circulation or Weddell Sea deep convection[22]. Critically, the four cores that do show a clear d_{ln} response collectively sample water vapour from all ocean basins (Extended Data Fig. 8a), suggesting that the changes to Southern Hemisphere atmospheric circulation are zonally coherent and involve all ocean basins (rather than just the Pacific basin, as demonstrated previously with WDC).

Figure 4a compares the two independent signals that we attribute to a change in atmospheric circulation: PC2 of the $\delta^{18}O$ response and the Antarctic mean d_{ln} response. Their time evolution is nearly identical, suggesting that they are distinct but consistent manifestations of the atmospheric circulation change. The SAM and Pacific–South American (PSA) pattern are the leading modes of large-scale Southern Hemisphere atmospheric variability with strong influence on Antarctic temperature[9,23]. We focus our analysis on East Antarctica, where we infer the largest response. The SAM (Fig. 4b) clearly impacts East Antarctic surface air temperature (SAT) strongly (correlation $|r|$ up to 0.65) and with the correct sign to explain the warming seen at EDC, DF and TAL. East Antarctic warming is seen for a more negative SAM index, driven primarily by anomalous atmospheric heat advection[24] (the observed cooling at EDML is discussed below). The PSA (Fig. 4c), on the other hand, is not meaningfully correlated with the SAT at the East Antarctic sites ($|r| \leq 0.15$). We further create a synthetic index that is the projection of the atmospheric loadings (Fig. 2h) onto the SAT anomaly inferred from the reanalysis at the core locations (Methods). The patterns in the SAT and the geopotential height associated with this index (Fig. 4d) closely resemble those of the SAM, with warming in East Antarctica and roughly annular geopotential height anomalies.

These tests suggest that the SAM is the closest present-day analogue to the temperature response that we identify in the ice-core record, corroborating our independent evidence from the d_{ln} data. Although
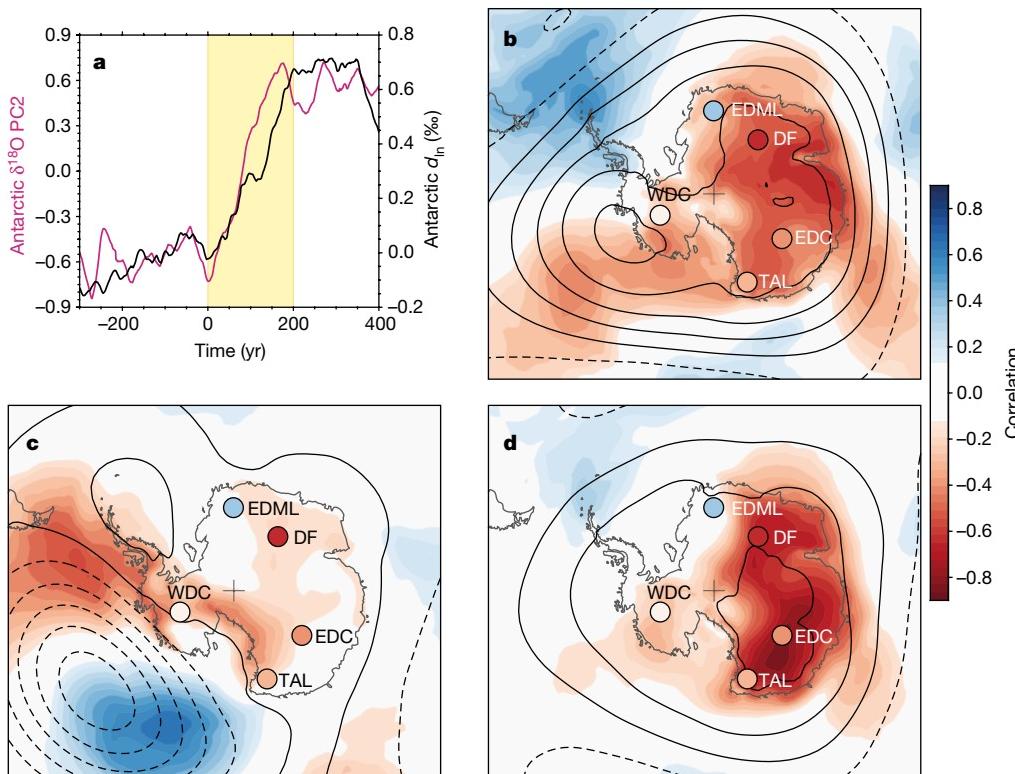
Fig. 4 | Attribution of the atmospheric mode of Antarctic temperature variability. a, Comparison of the principal component PC2 of the five Antarctic $\delta^{18}O$ stacks shown in Fig. 2d (pink; left axis) with the Antarctic mean d_{ln} stack shown in Fig. 3a (black; right axis). Isotope ratios are on the VSMOW scale. **b**, Correlation between a standardized monthly SAM index and the SAT (2-m temperature) in ERA-Interim[30] for 1979–2017 (shading; colour bar on the right) with superimposed 850-hPa geopotential height regressions (10-m contours) and the ice-core atmospheric temperature mode from Fig. 2h (circles; scale bar from Fig. 2). We note that regions of anti-correlation are coloured red (that is, warming in response to a negative SAM shift). **c**, As in **b**, but for a standardized PSA index. The SAM and PSA are taken to be the PC1 and PC2 principal components of the 850-hPa geopotential height field south of 20° S, respectively. **d**, As in **b**, but for a synthetic index of the atmospheric mode created by regressing ERA-Interim SAT anomalies at the ice-core sites onto the coefficients in Fig. 2h (Methods).

the PSA pattern may have been active during the DO cycle, it does not dominate the Antarctic response.

Our observation-based inferences on the timing and sign of changes to the Southern Hemisphere westerlies and the SAM are consistent with coupled atmosphere–ocean GCM simulations in which AMOC transitions are induced by North Atlantic freshwater forcing[6,12,25]. Such model simulations show a positive shift in the SAM index in response to AMOC shutdown and vice versa (Fig. 3a, b); this shift is synchronous with the applied forcing within uncertainty (Extended Data Table 1). Our observed atmospheric response is more gradual than the model-simulated SAM shift, possibly because of the (multi-decadal) data resolution in some cores and the fact that the d_{ln} signal is integrated over a large moisture source area extending to 20° S.

Next, we address differences between the ice-core data and the modern-day correlation pattern (Fig. 4b), most notably at EDML. The reanalysis correlation pattern captures the SAT response to monthly internal SAM variability, representing atmospheric heat advection anomalies[24]. The ice cores, on the other hand, record the response to a persistent long-term shift in the SAM[13,26], driving changes in SST, stratification and sea ice extent[22,26]. We speculate that on longer timescales the oceanic influence of the Weddell Sea drives the cooling at EDML owing to, for example, enhanced sea ice cover[22] and stratification and a weakening of the wind-driven Weddell Gyre. The negligible warming response at WDC is consistent with the relatively weak influence of the SAM in West Antarctic seen in the monthly reanalysis (Fig. 4b). Our observations may help to constrain the long-term response to a persistent SAM shift, on which GCMs disagree[13].

Lastly, we want to highlight additional structure in the Antarctic $\delta^{18}O$ stacks that is currently not part of scientific discourse on interhemispheric climate coupling. Most notably, Antarctic warming appears to

slow down around $t = -400$ yr (Fig. 2b), forming a secondary change point that precedes the abrupt DO warming events[27]. Likewise, the rate of Antarctic cooling appears to increase 200 years before the abrupt DO cooling events (Extended Data Fig. 7b). These secondary change points are subtler than the ones analysed in this work, have no apparent corresponding features in Antarctic d_{ln} or Greenland climate and remain unexplained.

In conclusion, our results show that Antarctica is influenced by abrupt changes in Northern Hemisphere climate on two distinct timescales, representing a slow oceanic and a fast atmospheric teleconnection. In particular, we provide observational evidence for zonally coherent meridional shifts in the Southern Hemisphere westerly winds in phase with Greenland DO events and their impact on Antarctic temperature. Such shifts have implications for global ocean circulation, Southern Ocean upwelling and productivity, and atmospheric CO_2[1–3]. It is therefore paramount to consider the DO cycle from a global, rather than a purely North Atlantic, perspective.

Online content

1. Marshall, J. & Speer, K. Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nat. Geosci.* **5**, 171–180 (2012).
2. Toggweiler, J. R., Russell, J. L. & Carson, S. R. Midlatitude westerlies, atmospheric CO_2, and climate change during the ice ages. *Paleoceanography* **21**, PA2005 (2006).

3.   Biastoch, A., Boning, C. W., Schwarzkopf, F. U. & Lutjeharms, J. R. E. Increase in Agulhas leakage due to poleward shift of Southern Hemisphere westerlies. *Nature* **462**, 495–498 (2009).
4.   Pritchard, H. D. et al. Antarctic ice-sheet loss driven by basal melting of ice shelves. *Nature* **484**, 502–505 (2012).
5.   Lee, S. Y., Chiang, J. C., Matsumoto, K. & Tokos, K. S. Southern Ocean wind response to North Atlantic cooling and the rise in atmospheric CO_2: modeling perspective and paleoceanographic implications. *Paleoceanography* **26**, PA1214 (2011).
6.   Pedro, J. B. et al. Beyond the bipolar seesaw: toward a process understanding of interhemispheric coupling. *Quat. Sci. Rev.* **192**, 27–46 (2018).
7.   Schneider, T., Bischoff, T. & Haug, G. H. Migrations and dynamics of the intertropical convergence zone. *Nature* **513**, 45–53 (2014).
8.   Markle, B. R. et al. Global atmospheric teleconnections during Dansgaard–Oeschger events. *Nat. Geosci.* **10**, 36–40 (2017).
9.   Ding, Q., Steig, E. J., Battisti, D. S. & Kuttel, M. Winter warming in West Antarctica caused by central tropical Pacific warming. *Nat. Geosci.* **4**, 398–403 (2011).
10. Buiron, D. et al. Regional imprints of millennial variability during the MIS 3 period around Antarctica. *Quat. Sci. Rev.* **48**, 99–112 (2012).
11. Stocker, T. F. & Johnsen, S. J. A minimum thermodynamic model for the bipolar seesaw. *Paleoceanography* **18**, 1087 (2003); correction **20**, PA1002 (2005).
12. Rind, D. et al. Effects of glacial meltwater in the GISS coupled atmosphere–ocean model. 2. A bipolar seesaw in Atlantic Deep Water production. *J. Geophys. Res. Atmos.* **106**, 27355–27365 (2001).
13. Kostov, Y. et al. Fast and slow responses of Southern Ocean sea surface temperature to SAM in coupled climate models. *Clim. Dyn.* **48**, 1595–1609 (2017).
14. Jouzel, J. et al. Validity of the temperature reconstruction from water isotopes in ice cores. *J. Geophys. Res.* **102**, 26471–26487 (1997).
15. WAIS Divide Project Members. Precise interpolar phasing of abrupt climate change during the last ice age. *Nature* **520**, 661–665 (2015).
16. Merlivat, L. & Jouzel, J. Global climatic interpretation of the deuterium–oxygen 18 relationship for precipitation. *J. Geophys. Res.* **84**, 5029–5033 (1979).
17. Masson-Delmotte, V. et al. GRIP deuterium excess reveals rapid and orbital-scale changes in Greenland moisture origin. *Science* **309**, 118–121 (2005).
18. Masson-Delmotte, V. et al. Abrupt change of Antarctic moisture origin at the end of Termination II. *Proc. Natl Acad. Sci. USA* **107**, 12091–12094 (2010).
19. Schmidt, G. A., LeGrande, A. N. & Hoffmann, G. Water isotope expressions of intrinsic and forced variability in a coupled ocean-atmosphere model. *J. Geophys. Res.* **112**, D10103 (2007).
20. Uemura, R. et al. Ranges of moisture-source temperature estimated from Antarctic ice cores stable isotope records over glacial–interglacial cycles. *Clim. Past* **8**, 1109–1125 (2012).
21. Sodemann, H. & Stohl, A. Asymmetries in the moisture origin of Antarctic precipitation. *Geophys. Res. Lett.* **36**, L22803 (2009).
22. Lefebvre, W., Goosse, H., Timmermann, R. & Fichefet, T. Influence of the Southern Annular Mode on the sea ice–ocean system. *J. Geophys. Res.* **109**, C09005 (2004).
23. Thompson, D. W. J. & Wallace, J. M. Annular modes in the extratropical circulation. Part I: month-to-month variability. *J. Clim.* **13**, 1000–1016 (2000).
24. Sen Gupta, A. & England, M. H. Coupled ocean-atmosphere-ice response to variations in the Southern Annular Mode. *J. Clim.* **19**, 4457–4486 (2006).
25. Liu, Z. et al. Transient simulation of last deglaciation with a new mechanism for Bolling–Allerod warming. *Science* **325**, 310–314 (2009).
26. Ferreira, D., Marshall, J., Bitz, C. M., Solomon, S. & Plumb, A. Antarctic Ocean and sea ice response to ozone depletion: a two-time-scale problem. *J. Clim.* **28**, 1206–1226 (2015).
27. Raisbeck, G. M. et al. An improved north–south synchronization of ice core records around the 41 kyr [10]Be peak. *Clim. Past* **13**, 217–229 (2017).
28. Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S. & Jouzel, J. Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
29. Rhodes, R. H. et al. Enhanced tropical methane production in response to iceberg discharge in the North Atlantic. *Science* **348**, 1016–1019 (2015).
30. Dee, D. P. et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).

Author contributions Data analysis by C.B., M. Severi, M. Sigl and J.J.W.; manuscript preparation by C.B.; volcanic ice-core synchronization by M. Sigl, M. Severi, C.B., J.R.M., F.P., S.F. and T.J.F.; GCM simulations and interpretation by F.H., J.B.P. and J.J.W.; moisture tagging/tracing experiments by B.R.M. and H.S.; ice-core water isotope analysis by K.G.-A., K.K., H.M., M.H., R.U., B.S. and E.J.S.; all authors discussed the results and contributed towards improving the final manuscript.

METHODS

Volcanic ice-core synchronization. Volcanic reference horizons provide the most precise way to synchronize ice-core age scales[31–35]. Within the last decade, progress has been made in volcanically linking the EDC ice core to the EDML core[31], the TAL core[32] and the DF core[33]. Here we provide new volcanic stratigraphic links between the WDC and the EDML, EDC and TAL cores using pattern matching of volcanic peaks in high-resolution records of either sulfur (WDC) or sulfate (EDML, EDC and TAL). Extended Data Fig. 1b summarizes the various synchronizations, with previously published ones indicated with grey arrows and the new ones presented here indicated with coloured arrows.

Volcanic synchronization via sulfur, sulfate or electrical conductivity measurement (ECM) records is a commonly used technique for Antarctic ice cores, and we rely on previously established methods described in detail elsewhere[31–35]. Matches are made by identifying sequences of sulfur peaks that have the same relative spacing in both cores[31–33,35]. Additional confidence in the match points comes from approximately proportional acid-concentration levels, smooth variations in the resulting annual layer thickness between stratigraphic tie-points, and in some cases a distinctive shape of the common signals in the different ice cores. We identify 773 volcanic ties between WDC and EDML, 396 between WDC and EDC, and 425 between WDC and TAL (Supplementary Data).

Stratigraphic matching is performed independently by two of the authors (analyst 1 and analyst 2) and then compiled and compared by a third author. Both analysts use an iterative approach, in which they first identify the major, unambiguous events. After marking these events (or clusters thereof), they re-align the datasets and replot them with the marked events overlapping. At this point, usually several of the smaller events are nearly on top of each other. These events are then marked and the data are replotted with the newly marked events overlapping, and the process is repeated.

We distinguish three cases: 'doubly', 'singly' and 'inconsistently' identified events. Doubly identified matches are cases where both analysts identify the same stratigraphic match between two cores for a given volcanic event (within a margin of a few centimetres). Singly identified matches are cases where only one of the two analysts identifies a stratigraphic match. Inconsistently identified matches are cases where a single volcanic peak in one core is linked to two different volcanic peaks in the other core. Around 99% of the singly identified events are found by analyst 1, demonstrating a difference in event detection threshold. For example, in the WDC–EDC synchronization the median volcanic peak sizes in WDC non-sea-salt sulfur are 67.7 and 31.4 p.p.b. for the doubly and singly identified matches, respectively, demonstrating that analyst 2 is more conservative in assigning match points. For comparison, the background non-sea-salt sulfur level is 15 ± 4 p.p.b.

Here, all the doubly identified events are assumed to be correct and retained. In the (relatively rare) case of an inconsistently matched event, the stratigraphic links suggested by both analysts are discarded to avoid ambiguity. Sequences of singly identified events are retained only if they occur in between two doubly identified match points. In the case of an inconsistently identified event (which is discarded), all singly identified matches adjacent to it are discarded also, until another doubly identified match is encountered. We give some examples below of hypothetical sequences of tie points, and how they are dealt with. In the following, 'd', 's' and 'i' denote doubly, singly and inconsistently identified tie points, respectively.

d–s–s–s–d. This is a hypothetical series of three **s** tie points between two **d** tie points. Because both analysts agree on the tie points on either end, there is no reason to assume that the **s** tie points are incorrect; it simply reflects the fact that analyst 2 is more conservative in assigning tie points than analyst 1. Therefore all tie points are retained in the final synchronization (**d–s–s–s–d**); **s** tie points are retained in such cases irrespective of how many are present in the series (for example, a series of ten **s** ties would be retained if bracketed on either side by a **d**).

d–i–d. In this case the **i** tie is removed, but the **d** ties are retained, resulting in the sequence **d–d** in the final synchronization.

d–s–s–i–s–d. The **i** tie is removed in all cases. However, in this example the **s** tie points occur adjacent to an **i** tie point, which casts doubt on their reliability. Therefore they are removed together with the inconsistent tie, and only the sequence **d–d** is retained in the final synchronization.

The matching is described here for the 10–61 kyr interval of interest (the WDC–EDC synchronization extends back only to 57 kyr BP). In synchronizing WDC and EDC, the two analysts identified 473 matches, out of which 103 were doubly, 8 were inconsistently and 362 were singly identified (all but two of which were identified by analyst 1). The final selection retained the 103 doubly identified matches and 293 of the singly identified matches; 69 singly identified matches were discarded because they were bracketed on either side by an inconsistent match. The 8 inconsistently identified matches all differed by less than 70 cm in EDC (or about 65 yr).

In synchronizing WDC and EDML, the two analysts identified a total of 793 matches, out of which 247 were doubly identified, 3 matches were inconsistently identified (all adjacent, and not separated by doubly identified events) and 543 were singly identified (all by analyst 1). The final selection retained the 247 doubly

identified matches and 526 of the singly identified matches; 17 singly identified matches were discarded because they were bracketed on either side by an inconsistent match. The 3 inconsistently identified matches all differed by less than 1.6 m in EDML (or about 65 yr).

TAL proved to be the most difficult of the cores to synchronize, presumably because the layers are more strongly compressed in this intermediate-depth core. The first attempt at synchronization yielded 55 doubly and 5 inconsistently identified matches, with most of the errors in the 770–905 m depth range (up to 5-m offsets in TAL). In light of this inconsistency, the analysts reviewed their volcanic ties throughout the core, with particular focus on the problematic interval. In the revised synchronization, the analysts identified a total of 437 matches, out of which 253 were doubly identified, 4 matches were inconsistently identified (all adjacent, and not separated by doubly identified events) and 180 were singly identified (all but 9 by analyst 1). The final selection retained the 253 doubly identified matches and 172 of the singly identified matches; 8 singly identified matches were discarded. The 4 inconsistently identified matches all differed by less than 75 cm in TAL (or about 75 yr). So while the final synchronization shows good agreement between the two analysts, we feel obliged to also report the first, less successful attempt. The TAL depth range that is hardest to synchronize to WDC (770–905 m TAL depth, or 14.9–25.2 kyr BP) also yielded no matches to EDC in a published study[32] (see yellow triangles at the bottom of Extended Data Fig. 1a). We note that this problematic time interval of 15–25 kyr BP does not influence the main results of this work—none of the abrupt DO events used in the stacking lie in this interval (DO event 2 is excluded from the stacking because the absence of an abrupt CH_4 signal precludes precise synchronization to Greenland[15]).

The number of doubly and inconsistently identified events provides one way to assess the reliability of the volcanic synchronization; the percentage of inconsistently identified events (out of the pool of doubly and inconsistently identified events) ranges from 1% to 7% for the three cores. Errors tend to occur in clusters of adjacent picks owing to the misidentification of a sequence of peaks; seen in this light, each of the WDC–EDML and WDC–TAL synchronizations contains only a single inconsistently identified sequence. The observed inconsistencies are always relatively small and of the order of a few decades. A second method of assessing the reliability of the volcanic synchronization is via the redundancy offered by having multiple cores. Whenever three ice cores in Extended Data Fig. 1b are connected via three independent synchronizations (that is, whenever the arrows form a triangle), this offers the possibility to test the internal consistency of the synchronization. Over the age interval of interest (the last 61 kyr), 213 ties had been previously identified between EDML and EDC[36], as well as 102 ties between TAL and EDC. This introduces a degree of redundancy that allows testing the internal consistency of the synchronization. EDC, TAL and EDML are volcanically synchronized within the AICC2012 (Antarctic Ice Core Chronology)[36], whereas WDC uses the independent WD2014 timescale[37,38]. For each volcanic tie point, the difference between the WD2014 age and the AICC2012 age is shown in Extended Data Fig. 1a. EDC and EDML are volcanically synchronized over the last 60 kyr (blue triangles), and therefore the excellent agreement between the blue (WDC age minus EDML age) and red (WDC age minus EDC age) curves in Extended Data Fig. 1a shows the volcanic framework to be internally consistent. Likewise, for the period 25–42 kyr BP and <13 kyr BP, where TAL and EDC are volcanically synchronized (yellow triangles), the yellow (WDC age minus TAL age) and red curves agree well, suggesting internal consistency. Beyond 42 kyr BP the yellow curve deviates from the blue and red ones, suggesting that the TAL core is imperfectly synchronized within the AICC2012 chronology (owing to an absence of volcanic ties at the time).

All ice cores used in this study were synchronized to the WD2014 chronology[37,38]. For the four non-WDC cores, we start from their original ice-core chronologies; this is the AICC2012 chronology[36] for the EDML, EDC and TAL ice cores, and the DF2006 chronology for the DF core[39]. For each core we add a time-variable offset to the WD2014 chronology that is obtained using linear interpolation between the volcanic tie points. For the EDML, EDC and TAL cores we have direct synchronizations to WDC as described above. For the DF core, synchronization is indirect via the EDC core (Extended Data Fig. 1b). Previously, 297 tie points had been identified[33] between EDC and DF in the interval of interest (mean spacing of 173 years). These volcanic ties are transferred to WDC using the WDC–EDC synchronization described above.

Over the time interval of interest, the offset of the AICC2012 cores (EDML, EDC and TAL) ranges roughly from −330 yr to +430 yr, with an average offset of −10 yr (with negative values meaning that WD2014 is younger than AICC2012). For the DF core the range is from −230 to 1,884 yr, with an average of +739 yr.

Uncertainty in volcanic synchronization. There are two types of uncertainty to consider. First, the volcanic ties themselves may be incorrect; second, between ties the interpolation strategy introduces an error. The first type is difficult to quantify. Either the ties are correctly identified and the relative age uncertainty is zero, or the ties are false and the relative age uncertainty is infinite (that is, we have learned

nothing). Past studies have sometimes assigned Gaussian errors to volcanic tie points; although this is a practical necessity for applications that optimize the fit to multiple age constraints[36,40–42], it does not reflect the true uncertainty meaningfully and is not applied here.

We have high confidence in the correctness of the volcanic ties because of the internal consistency of the new volcanic ties with previously published ones (Extended Data Fig. 1a), and the fact that doubly identified ties greatly outnumber the inconsistently identified ones. We have removed inconsistent matches from the synchronization, and we assume the remaining matches to be correct.

The second type of uncertainty is due to interpolation between volcanic ties. This introduces an age uncertainty that depends on the age difference between adjacent ties, L. We estimate the interpolation uncertainty using the layer-counted section of the WD2014 chronology, which goes back to 31.2 kyr BP. To estimate the interpolation uncertainty for two volcanic markers that are, say, $L = 100$ yr apart, we can randomly pick thousands of 100-yr intervals from the WD2014 chronology. Within each of these, the age evolution deviates from the assumed linear interpolation, and 1σ of this age deviation is used as the interpolation uncertainty. Typical results are shown in Extended Data Fig. 2a for several values of L. We find that the maximum uncertainty scales as $\propto L^2$. Compared with East Antarctica, West Antarctica receives a larger contribution of its snowfall from storm systems and synoptic activity, making accumulation rates more variable[43,44]; the estimates given here should thus be considered conservative when used in the interior of Antarctica. The volcanic interpolation uncertainty for the four cores is plotted in Extended Data Fig. 2b. For DF, synchronization to WDC is done using EDC as an intermediary core, and therefore the two synchronization errors are added in quadrature.

In our synchronization we use both the singly and doubly identified tie points and treat them equally. We acknowledge, however, that the doubly identified ties are more reliable. Therefore, we repeat the analyses described in the main text using only the doubly identified tie points (as opposed to both singly and doubly identified tie points) and we find that the conclusions of this work do not depend on this choice of tie points. Those tests are not shown here, but alternative chronologies that use only the doubly identified tie points and alternative versions of the manuscript figures showing those analyses are available from the corresponding author upon request.

Ice-core water stable-isotope data. A combination of previously published and unpublished ice-core water isotope data ($\delta^{18}O$ and δD; where δD represents the $^2H/^1H$ composition) are used in this study. The deuterium excess (d_{ln}) is calculated from the $\delta^{18}O$ and δD isotope ratios using the logarithmic definition of ref. [20]:

$$d_{ln} = \ln(1 + \delta D) - 8.47 \ln(1 + \delta^{18}O) + 0.0285 [\ln(1 + \delta^{18}O)]^2$$

For WDC we use previously published $\delta^{18}O$ and δD data[8,15,45] measured using laser spectroscopy. The data have a typical depth resolution of 1 m for the 0–2.3 kyr BP interval, of 0.5 m for 2.3–56 kyr BP and of 0.25 m for 56–68 kyr BP; this corresponds to an average time resolution of 17 yr for the study period (11–61 kyr BP). Using the centimetre-scale continuous-flow-analysis record of WDC $\delta^{18}O$ instead gives identical results to those presented here[46].

For EDML we use previously published $\delta^{18}O$ and δD data[47,48] measured using conventional isotope ratio mass spectrometry (IRMS). The data have a typical depth resolution of 0.5 m, corresponding to an average time resolution of 24 yr for the study period.

For EDC we use previously published $\delta^{18}O$ and δD data[48,49] measured using conventional IRMS. The data have a typical depth resolution of 0.55 m, corresponding to an average time resolution of 44 yr for the study period.

For DF we use new and published water isotope data[39,50,51]. Two datasets are used. The first contains $\delta^{18}O$ data measured using IRMS in the 300–1,151 m depth range (10–71 kyr BP) at 0.5-m resolution[50]. The second is a set of $\delta^{18}O$ and δD data measured using IRMS in the 550–849 m depth range (23–45 kyr BP); this dataset was measured at 0.1-m resolution and averaged into 0.5-m bins. We note that d_{ln} is only available from the second dataset, which spans DO events 2–11 (DO event numbering is shown in Fig. 1). In the depth range of overlap, $\delta^{18}O$ data from both datasets are averaged (after correcting the second dataset by $+0.213‰$ to account for a calibration offset) to produce the final time series. The combined $\delta^{18}O$ record has an average time resolution of 35 yr for the study period.

For TAL we use a combination of new and previously published[10,52,53] data. Bag-average, 1-m resolution $\delta^{18}O$ and δD data measured using IRMS are available for the entire core. High-resolution (0.1 m) $\delta^{18}O$ data measured using IRMS are available in the 598–786 m (10–16 kyr BP) and 1,030–1,282 m (37–65 kyr BP) depth ranges. High-resolution $\delta^{18}O$ data are averaged into 0.5-m bins and combined with bag-average data for the remaining depth intervals. The combined $\delta^{18}O$ record has an average time resolution of 50 yr for the study period.

Stacking procedure. The stacking procedure used to investigate the Antarctic climate response to abrupt DO variability is described in detail elsewhere[15].

In short, the individual Greenland events are aligned at the midpoint of their abrupt $\delta^{18}O$ transitions (either DO warming or DO cooling). All Antarctic events (on their volcanically synchronized WD2014 timescales) are aligned at the midpoints of the abrupt WDC CH_4 transitions. We then average over events to obtain the shared climatic signal; to derive the north–south phasing we use the independently established CH_4 delay of 56 ± 19 yr (1σ) behind $\delta^{18}O$ in Greenland[54].

For one of the abrupt events (DO 10), improved inter-polar synchronization data are available from ^{10}Be variations during the Laschamp event (41 kyr BP) between the Greenland NGRIP core and the Antarctic EDC and EDML cores[27]; these timing constraints are incorporated into our stacking procedure (DO 10 only). The EDML, EDC, DF and TAL ice cores have CH_4 records with much higher (gas age−ice age) difference (Δage) uncertainty and lower resolution than WDC, therefore the north–south phasing precision cannot be improved by considering CH_4 synchronization for those cores also.

In this work we consider DO 0 (that is, the Younger Dryas–Holocene transition) through DO 16; DO 17 falls outside the volcanic synchronization for the EDC and DF cores. DO 2 is omitted owing to the absence of a clear CH_4 response, precluding synchronization. All stacked records are shown in Extended Data Fig. 3.

Uncertainty in the timing of the stacked records comes from the following sources: (1) the Δage in the WDC core[37]; (2) DO midpoint detection in the abrupt NGRIP $\delta^{18}O$ record; (3) DO midpoint detection in the abrupt WDC CH_4 record; (4) interpolation of the WDC chronology between tie points; (5) the climatic lag of 56 ± 19 yr (1σ) of atmospheric CH_4 behind Greenland $\delta^{18}O$[54]; (6) trend analysis in the BREAKFIT[55] and RAMPFIT[56] fitting routines; and (7) volcanic synchronization to the WD2014 chronology.

The combined uncertainty due to the first five items was assessed previously using a Monte Carlo routine, suggesting a 1σ timing uncertainty of 38 yr and 41 yr in the WDC stacks for DO warming and DO cooling, respectively[15]. The uncertainty in the trend analysis is given in Extended Data Table 1. The uncertainty in the volcanic synchronization is shown in Extended Data Fig. 2; averaged over the stacked events, the 1σ synchronization uncertainty is 0 yr at WDC, 2 yr at EDML, 6 yr at EDC, 13 yr at DF and 2 yr at TAL.

By stacking only the most prominent DO events (those following Heinrich events; that is, DO events 0, 1, 4, 8, 12 and 14) or just the minor ones (the remainder), we find that the magnitude of the atmospherically forced Antarctic response is larger for the former, suggesting proportionality with the climate perturbation (Extended Data Fig. 4a–f). Proportionality of the atmospheric response is further seen for individual events (Extended Data Fig. 4g; see the figure legend for details).

PCA. PCA allows different climatic modes to be identified in (palaeoclimatic) time series from different locations[57]. Here we perform PCA on the stacked $\delta^{18}O$ records at the five individual sites using the MATLAB function *pca*. The $\delta^{18}O$ stacks discussed in the main text combine 19 individual events—all those that fall within the volcanic synchronization interval. To assess the sensitivity of our conclusions to including or excluding individual events, we perform additional experiments in which we stack only n events (rather than all 19). The n events are randomly sampled without replacement (that is, any given event cannot be picked twice for each stacking). We then perform PCA of these stacked records (the same events are stacked for each of the cores) and standardize the principal component vectors by taking the z-score (or standard score). Extended Data Fig. 5 shows typical results for $n = 2$ and $n = 8$, where for each n we repeat the experiment 50,000 times to obtain reliable statistics. Extended Data Fig. 5a, b shows PC1 and PC2, respectively, with the solid line showing the mean of 50,000 experiments and the shaded envelope the associated $\pm 1\sigma$. Extended Data Fig. 5c shows a histogram of the percentage of variance explained by each of the modes.

We find that even by stacking as few as just $n = 2$ events, the method can, on average, identify the oceanic and atmospheric components described in the main text. Perhaps not surprisingly, we find that when including fewer events the signal-to-noise ratio decreases: with fewer events the estimated signal amplitude is smaller in both PC1 and PC2, and the uncertainty envelope is wider. As more events are included in the stacking, the percentage of variance explained by PC1 increases as the coherence between the various Antarctic cores increases owing to the improved signal-to-noise ratio.

The analysis discussed in the main text uses a 1,500-yr window (centred around $t = 100$ yr), which is chosen because it corresponds to the recurrence time of the shortest DO cycles[58–61]. The variance explained by the oceanic and atmospheric modes depends on the window length, as shown in Extended Data Fig. 6a. For window lengths exceeding about 750 yr, the oceanic mode explains most of the variance (PC1), with the atmospheric mode explaining less. However, at short window lengths (<750 yr) the atmospheric mode explains most of the variance, making it PC1. The comparison of PC1 at a 400-yr window with PC2 at a 2,000-yr window (Extended Data Fig. 6b) illustrates the ability of the PCA to identify the atmospheric mode at different window lengths. The crossover behaviour (that is, the atmospheric mode shifts from being PC2 to being PC1 as a function of window length) is due to the fact that the signal variance of the step-like atmospheric mode

occurs chiefly within the $t = 0$ to $t = 100$ yr interval; signal variance of the oceanic mode depends strongly on the window length owing to its gradual nature.

Lag-time analysis of PC1 and PC2 using the RAMPFIT[56] and BREAKFIT[55] routines confirms the crossover behaviour. At short window lengths (<700 yr), PC1 is characterized by the instantaneous or fast response of the hypothesized atmospheric teleconnection, whereas at large window lengths (>800 yr) it shows the hypothesized 200-yr delayed oceanic teleconnection. PC2 shows the opposite behaviour; however, we note that for PC2 at a 400-yr window length no meaningful solution can be found with either routine. At window lengths >700 yr the lag times remain stable and vary only within the uncertainty bound. Unless specified otherwise, a 1,500-yr window is used in this work.

Robustness of the atmospheric spatial pattern. The spatial pattern that we identify for the atmospheric mode is one of the main results of this work, and we here test its robustness. In Extended Data Fig. 6d–f we compare three alternative ways of identifying the spatial pattern; the Pearson correlation coefficient (r) between the shown pattern and the pattern identified in the main text (EOF2 at a 1,500-yr window) is given for each. Details on the three methods are given in the figure legend. Correlation coefficients ranging from $r = 0.92$ to $r = 0.999$ suggest that identification of the atmospheric pattern is both qualitatively and quantitatively robust.

Rotated PCA vectors. PCA aims to explain the largest amount of variance, whereas our goal is to distinguish between the oceanic and atmospheric modes. While PC1 is clearly dominated by the 200-yr-delayed oceanic bipolar seesaw (Fig. 2d), it appears that PC1 also captures some abrupt warming around $t = 0$, apparently because atmospherically induced warming is more prevalent over Antarctica than cooling (Northern Hemisphere DO warming case). We therefore construct (admittedly somewhat subjective) oceanic and atmospheric response functions (Fig. 2e), which are derived from the principal components in the following way. Let *PC1* and *PC2* be the first two principal components through time; these vectors are perpendicular. We let the atmospheric response, *ATM*, be identical to *PC2*. The oceanic response, *OCE*, is found by rotating the *PC1* vector in the plane spanned by vectors *PC1* and *PC2* over an angle of $-13°$:

$$OCE = \cos(-13°)PC1 + \sin(-13°)PC2$$

The rotation angle of $-13°$ is picked such that the $\delta^{18}O$ shift at $t = 0$ is minimized. The spatial pattern associated with *ATM* (Fig. 2h) is found by multiple linear regression of the $\delta^{18}O$ stacks at the individual sites to *OCE* and *ATM* using the MATLAB function *regress*.

SAM, PSA and synthetic 'atmospheric' indices. The SAM and PSA indices are calculated from climate-model and reanalysis data as the first and second mode of variability, respectively, in the PCA/EOF analysis (MATLAB function *pca*) of the sea-level pressure (model) or the 850-hPa geopotential height (reanalysis) south of 20° S, after subtracting the long-term mean and scaling the anomalies by the square root of the cosine of the latitude to account for the decreased surface area closer to the poles. A synthetic index of the inferred atmospheric teleconnection is created by projecting the reanalysis SAT anomalies at the ice-core locations onto the atmospheric loading coefficients from Fig. 2h, the spatial signature of which is shown in Fig. 4d. Mathematically, this projection is done by multiplying the SAT anomalies with the loading coefficients and summing over them at each monthly time step.

It is worth noting that all these inferences with respect to reanalysis data rely upon just five atmospheric loading coefficients from the limited (from the perspective of large-scale circulation) spatial domain of Antarctica, so it is difficult to rigorously exclude non-SAM atmospheric influences from the temperature pattern alone.

In Extended Data Fig. 9 we compare the variability in the Southern Hemisphere westerly winds calculated with glacial CCSM3 climate simulations[25,62] with that obtained from the ERA-Interim reanalysis[30]. Both results show greater variability in the Indian and Pacific sectors than in the Atlantic sector. Compared to ERA-Interim, the CCSM3 SAM is more zonal/annular in its structure; the CCSM3 westerlies also appear to have smaller variability (some of this difference could be due to comparing annual mean with decadal mean data). In CCSM3 the forced response of the Southern Hemisphere westerlies (at 19 kyr BP, in response to increased North Atlantic freshwater; right panel) is very similar to the internal variability of the Southern Hemisphere westerlies (middle panel), suggesting that the change in the Southern Hemisphere atmospheric circulation induced by freshwater forcing in the North Atlantic is analogous to the existing mode of internal variability.

To estimate the magnitude of SAM shifts of the DO cycle, we analyse the changes in central East Antarctica, where the signal is the largest and most consistent with the observed present-day relationship (Fig. 4). Using an isotope sensitivity of 0.8‰ K^{-1}, the atmospherically induced temperature anomaly in central East Antarctica (DF, EDC) is in the range 0.20–0.45 °C during DO warming (the lower and upper bounds reflect typical minor and major DO events, respectively). Regression of the ERA-Interim 2-m temperature at DF and EDC to the SAM index

shows a slope of around -1.2 °C per unit of normalized SAM (the normalized SAM index time series has a standard deviation of 1), implying a shift in the SAM index of around 0.2 to 0.4 normalized (modern-day) SAM units (rounded to one significant figure). This estimate assumes: (1) a linear isotope–temperature response using the modern-day spatial slope, and (2) that the monthly SAM–SAT regression from the monthly internal SAM variability also applies to persistent SAM shifts during the glacial DO cycle; both assumptions are subject to uncertainties that we do not address here. The CCSM3 model simulates a persisting SAM shift of the same magnitude as the internal SAM variability in decadal averaged data (Fig. 3); because internal SAM variability will be larger in monthly than in decadally averaged data, the model and database estimates may be in agreement. The reanalysis time period is too short to derive robust estimates of decadally averaged SAM variability.

GCM simulations. We analyse previously published TraCE-21k transient climate model simulations with CCSM3[25,62–65]. AMOC collapse and resumption are triggered using freshwater forcing in the North Atlantic. The 'DGL19k' run is used for the AMOC collapse and the 'DGL-overshoot-C' run for the AMOC resumption case[64].

Moisture origin analysis. We use two separate experiments to trace the moisture origin of the precipitation at the coring sites. The first method uses a Lagrangian moisture source diagnostic[66] based on a previously published dataset[67]. Using the winds, temperature and humidity of the ERA-Interim reanalysis dataset[30] covering the years 1980–2013, we calculate 5 million air parcel trajectories covering the global atmosphere at a resolution of 6 h using the Lagrangian particle dispersion model FLEXPART[68]. From the analysis of specific humidity changes over time along the air parcel trajectories, moisture sources are identified whenever the specific humidity in the air parcels increases more than a threshold value of 0.1 g kg^{-1} over a 6-h period. The fractional contribution of each moisture source to the final precipitation at the target location (an area of 300×300 km^2 centred on each ice-core site) is obtained from calculating the amount contributed by a moisture source to the humidity already present in an air parcel. For precipitation en route, the contributions of previous sources are proportionally discounted. This results in a quantitative estimate of the contribution of surface areas to the precipitation in the target region in units of evaporation (water mass per unit area per unit time), including their position (latitude and longitude). The moisture source contributions for each site and precipitation event are combined into mass-weighted annual mean values and scaled with respect to the total amount of water deposited at the target region.

The second method uses water tagging in a 50-yr simulation with the Community Atmosphere Model (CAM) with prescribed seasonally varying SST and modern boundary conditions. The experiment is set up to evaluate the meridional moisture source distribution, with further details and figures given in ref. [8]. In short, evaporation is tagged in 11 bins. One bin is the Antarctic continent (re-evaporation) and the ocean and ice shelves south of 70° S; the other ten bins are zonal bins of 5° latitude each, ranging from 20° S to 70° S. For each core, the moisture source distribution is found by evaluating the relative contributions from each of the bins for the last 30 years of the run. Two moisture source distributions are created, one for all years in which the mean annual SAM index was positive, and one for all years in which it was negative (Extended Data Fig. 8b).

Change-point detection. We use two well documented and widely used change-point detection methods, BREAKFIT[55] and RAMPFIT[56], with the results given in Extended Data Table 1. The choice of method is based on the shape of the time series $x(t)$. BREAKFIT finds a single change point and fits a linear slope on either side; these features make it suitable for the oceanic mode/PC1 discussed in the main manuscript. RAMPFIT finds two change points; it assumes that $x(t)$ has a constant value x_1 for $t < t_1$, then is ramped up or down until it reaches value x_2 at time t_2, after which it remains constant at value x_2 for $t > t_2$. These features make RAMPFIT suitable for the atmospheric mode/PC2 discussed in the main manuscript. To find the two change points in the $d_{\ln}$ stacks, we apply both the RAMPFIT and the BREAKFIT algorithms twice (once for each change point). The results are comparable (Extended Data Table 1), and in the main text we report the average of the two methods.

Code availability. The MATLAB code used for the stacking procedure can be found in the supplementary information of ref. [15] and is available from the corresponding author upon request.

Data availability

Source Data (WDC sulfur data, volcanic tie points and water isotope data on synchronized chronologies) and derived products (stacks, PCA results, etc.) are available in the online version of the paper and in the NOAA palaeoclimate data archive.

31. Severi, M. et al. Synchronisation of the EDML and EDC ice cores for the last 52 kyr by volcanic signature matching. *Clim. Past* **3**, 367–374 (2007).

32. Severi, M., Udisti, R., Becagli, S., Stenni, B. & Traversi, R. Volcanic synchronisation of the EPICA-DC and TALDICE ice cores for the last 42 kyr BP. *Clim. Past* **8**, 509–517 (2012).
33. Fujita, S., Parrenin, F., Severi, M., Motoyama, H. & Wolff, E. W. Volcanic synchronization of Dome Fuji and Dome C Antarctic deep ice cores over the past 216 kyr. *Clim. Past* **11**, 1395–1416 (2015).
34. Sigl, M. et al. Insights from Antarctica on volcanic forcing during the Common Era. *Nat. Clim. Chang.* **4**, 693–697 (2014).
35. Parrenin, F. et al. Volcanic synchronisation between the EPICA Dome C and Vostok ice cores (Antarctica) 0–145 kyr BP. *Clim. Past* **8**, 1031–1045 (2012).
36. Veres, D. et al. The Antarctic ice core chronology (AICC2012): an optimized multi-parameter and multi-site dating approach for the last 120 thousand years. *Clim. Past* **9**, 1733–1748 (2013).
37. Buizert, C. et al. The WAIS Divide deep ice core WD2014 chronology – part 1: methane synchronization (68–31 ka BP) and the gas age–ice age difference. *Clim. Past* **11**, 153–173 (2015).
38. Sigl, M. et al. The WAIS Divide deep ice core WD2014 chronology – part 2: annual-layer counting (0–31 ka BP). *Clim. Past* **12**, 769–786 (2016).
39. Kawamura, K. et al. Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* **448**, 912–916 (2007).
40. Bazin, L. et al. An optimized multi-proxy, multi-site Antarctic ice and gas orbital chronology (AICC2012): 120–800 ka. *Clim. Past* **9**, 1715–1731 (2013).
41. Parrenin, F. et al. IceChrono1: a probabilistic model to compute a common and optimal chronology for several ice cores. *Geosci. Model Dev.* **8**, 1473–1492 (2015).
42. Lemieux-Dudon, B. et al. Consistent dating for Antarctic and Greenland ice cores. *Quat. Sci. Rev.* **29**, 8–20 (2010).
43. Fudge, T. J. et al. Variable relationship between accumulation and temperature in West Antarctica for the past 31,000 years. *Geophys. Res. Lett.* **43**, 3795–3803 (2016).
44. Fortuin, J. & Oerlemans, J. Parameterization of the annual surface temperature and mass balance of Antarctica. *Ann. Glaciol.* **14**, 78–84 (1990).
45. WAIS Divide Project Members. Onset of deglacial warming in West Antarctica driven by local orbital forcing. *Nature* **500**, 440–444 (2013).
46. Jones, T. R. et al. Water isotope diffusion in the WAIS Divide ice core during the Holocene and last glacial. *J. Geophys. Res. Earth* **122**, 290–309 (2017).
47. EPICA Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444**, 195–198 (2006).
48. Stenni, B. et al. The deuterium excess records of EPICA Dome C and Dronning Maud Land ice cores (East Antarctica). *Quat. Sci. Rev.* **29**, 146–159 (2010).
49. Jouzel, J. et al. Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317**, 793–796 (2007).
50. Kawamura, K. et al. State dependence of climatic instability over the past 720,000 years from Antarctic ice cores and climate modeling. *Sci. Adv.* **3**, (2017).
51. Uemura, R. et al. Asynchrony between Antarctic temperature and CO_2 associated with obliquity over the past 720,000 years. *Nat. Commun.* **9**, 961 (2018).
52. Stenni, B. et al. Expression of the bipolar see-saw in Antarctic climate records during the last deglaciation. *Nat. Geosci.* **4**, 46–49 (2011).
53. Landais, A. et al. A review of the bipolar see–saw from synchronized and high resolution ice core water stable isotope records from Greenland and East Antarctica. *Quat. Sci. Rev.* **114**, 18–32 (2015).
54. Baumgartner, M. et al. NGRIP CH_4 concentration from 120 to 10 kyr before present and its relation to $\delta^{15}N$ temperature reconstruction from the same ice core. *Clim. Past* **10**, 903–920 (2014).
55. Mudelsee, M. Break function regression. *Eur. Phys. J. Spec. Top.* **174**, 49–63 (2009).
56. Mudelsee, M. Ramp function regression: a tool for quantifying climate transitions. *Comput. Geosci.* **26**, 293–307 (2000).
57. Clark, P. U., Pisias, N. G., Stocker, T. F. & Weaver, A. J. The role of the thermohaline circulation in abrupt climate change. *Nature* **415**, 863–869 (2002).
58. Grootes, P. M. & Stuiver, M. Oxygen 18/16 variability in Greenland snow and ice with 10^{-3}-to-10^5-year time resolution. *J. Geophys. Res.* **102**, 26455–26470 (1997).
59. Alley, R. B., Anandakrishnan, S. & Jung, P. Stochastic resonance in the North Atlantic. *Paleoceanography* **16**, 190–198 (2001).
60. Buizert, C. & Schmittner, A. Southern Ocean control of glacial AMOC stability and Dansgaard–Oeschger interstadial duration. *Paleoceanography* **30**, 1595–1612 (2015).
61. Ditlevsen, P. D., Kristensen, M. S. & Andersen, K. K. The recurrence time of Dansgaard–Oeschger events and limits on the possible periodic component. *J. Clim.* **18**, 2594–2603 (2005).
62. He, F. et al. Northern Hemisphere forcing of Southern Hemisphere climate during the last deglaciation. *Nature* **494**, 81–85 (2013).
63. Liu, Z. et al. Younger Dryas cooling and the Greenland climate response to CO_2. *Proc. Natl Acad. Sci.* **109** 11101–11104 (2012).
64. He, F. Simulating Transient Climate Evolution of the Last Deglaciation with CCSM3. PhD thesis, Univ. Wisconsin-Madison (2010).
65. Pedro, J. B. et al. The spatial extent and dynamics of the Antarctic Cold Reversal. *Nat. Geosci.* **9**, 51–55 (2016).
66. Sodemann, H., Schwierz, C. & Wernli, H. Interannual variability of Greenland winter precipitation sources: Lagrangian moisture diagnostic and North Atlantic Oscillation influence. *J. Geophys. Res.* **113**, D03107 (2008).
67. Läderach, A. & Sodemann, H. A revised picture of the atmospheric moisture residence time. *Geophys. Res. Lett.* **43**, 924–933 (2016).
68. Stohl, A., Forster, C., Frank, A., Seibert, P. & Wotawa, G. Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2. *Atmos. Chem. Phys.* **5**, 2461–2474 (2005).
69. Locarnini, R. A. et al. *World Ocean Atlas 2013, Volume 1: Temperature* (eds Levitus, S. & Mishonov, A.) (NOAA/NESDIS, Silver Spring, 2013).
70. Kalnay, E. et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
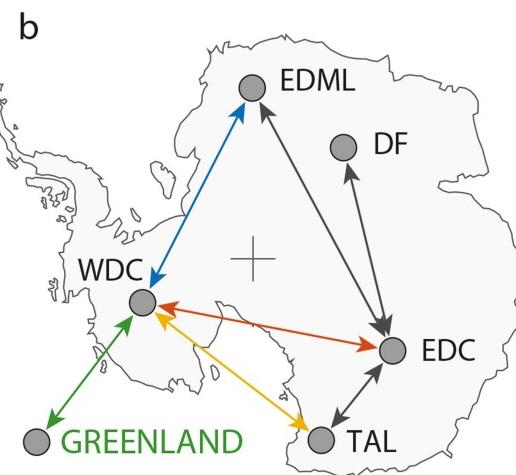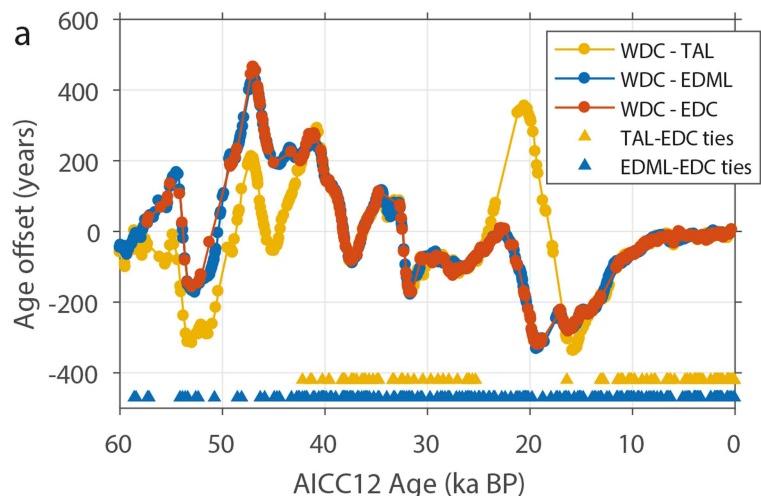
Extended Data Fig. 1 | Volcanic synchronization of Antarctic ice cores. **a**, Age offset between the WD2014[37,38] (WDC) and AICC2012[36] (TAL, EDML, EDC) age scales, with each dot representing a volcanic tie point. Yellow and blue triangles denote the timing of TAL–EDC and EMDL–EDC volcanic ties[32,36], respectively. **b**, Overview of synchronizations between the ice cores used in this study. Grey arrows indicate previously published synchronizations and coloured arrows denote synchronizations performed here. Synchronizations within Antarctica are based purely on volcanic links; synchronization between WDC and NGRIP (Greenland) are based on atmospheric CH_4 (green arrow).
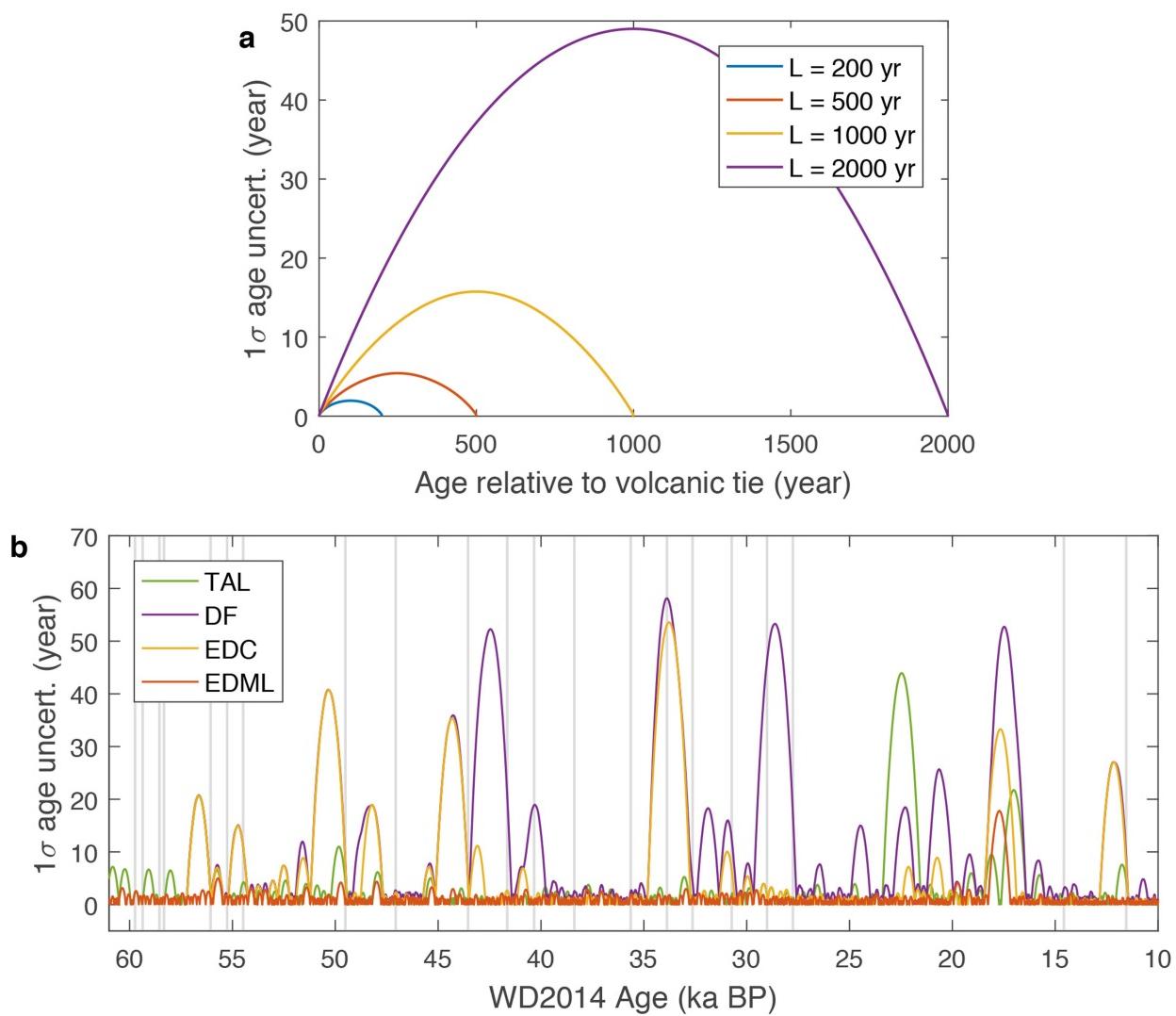
Extended Data Fig. 2 | Age uncertainty due to volcanic synchronization. **a**, Interpolation uncertainty (1σ) for four different values of L (the spacing between two adjacent volcanic tie points), based on the layer-counted WD2014 age scale[38]. **b**, Interpolation uncertainty in synchronizing the four ice cores to WD2014 chronology. Grey vertical lines give the timing of DO events.
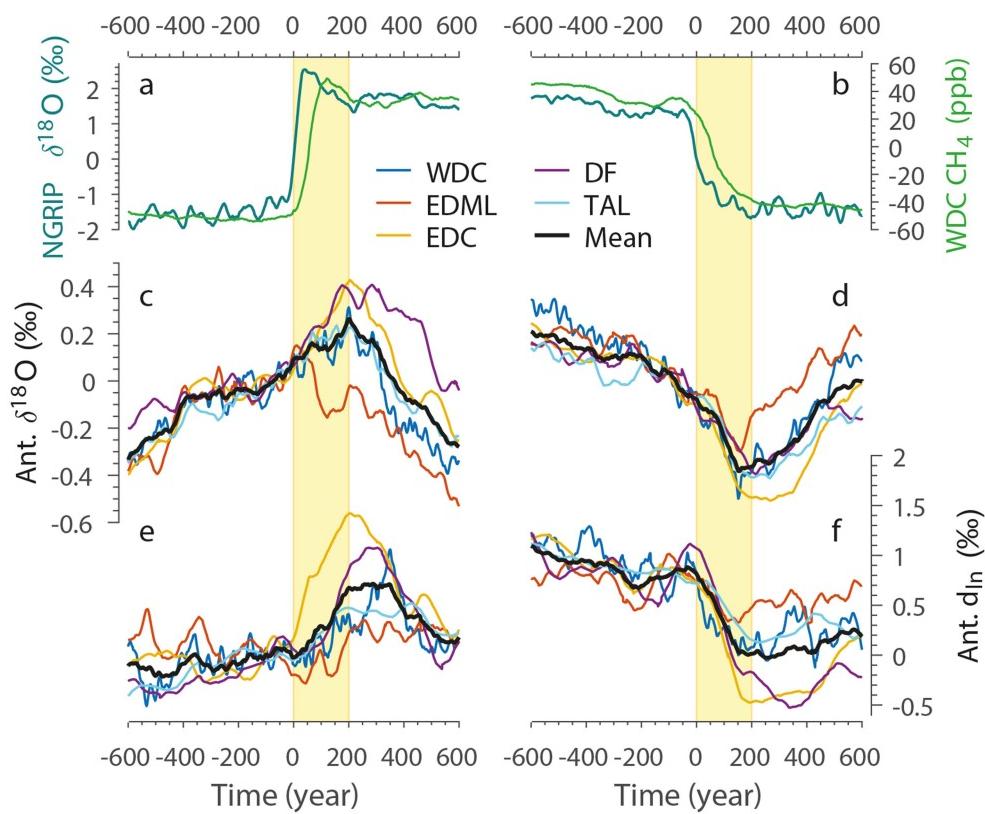
Extended Data Fig. 3 | Site-specific stacks of $\delta^{18}O$ and d_{ln}. a, Stack of NGRIP $\delta^{18}O$ (teal; left axis) and WDC CH_4 (green; right axis) during DO warming. **b**, As in **a**, but for DO cooling. **c**, Stack of Antarctic $\delta^{18}O$ at the indicated locations (see key) during DO warming. **d**, As in **c**, but for DO cooling. **e**, Stack of Antarctic d_{ln} at the indicated locations during DO warming. **f**, As in **e**, but for DO cooling. Isotope ratios are on the VSMOW scale.

Extended Data Fig. 4 | Proportionality of the atmospheric response. a–f, Comparison of stacks of just the major DO events (those following Heinrich events, namely, DO events 0, 1, 4, 8, 12 and 14; a, c, e) and just the minor DO events (the remainder; b, d, f). a, b, Stacks of NGRIP $\delta^{18}O$ (left axis) and CH_4 (right axis). c, d, Stacks of Antarctic $\delta^{18}O$ at the indicated locations. e, f, As in c and d, but with the Antarctic mean subtracted. g, Proportionality of the atmospheric response for individual events (numbered). The NGRIP event size is found via regression of individual NGRIP events to the multi-event NGRIP $\delta^{18}O$ stack normalized to unit variance (Fig. 2a). The Antarctic atmospheric response is found via multiple linear regression of single-site individual events to the atmospheric and oceanic modes (Fig. 2e). Shown are the average (dots) and standard deviation (error bars) of the response at EDC, DF and EDML (the cores with the strongest atmospheric response); the EDML response is multiplied by −1 because it has the opposite sign of the response at DF and EDC. Red and blue dots denote the major and minor DO events, respectively. Isotope ratios are on the VSMOW scale.

Extended Data Fig. 5 | Reduction of the number of events in the $\delta^{18}O$ stacks. a, PC1, when stacking $n=2$ and $n=8$ randomly selected events. The thick line and shaded area represent the mean and $\pm 1\sigma$, respectively, obtained from 50,000 runs. The vertical yellow bands denote the 200-yr period after the abrupt DO event at $t=0$. **b**, As in **a**, but for PC2. **c**, Fraction of signal variance explained by PC1 and PC2, when stacking 2 or 8 randomly selected events. Colour coding as in **a** and **b**.
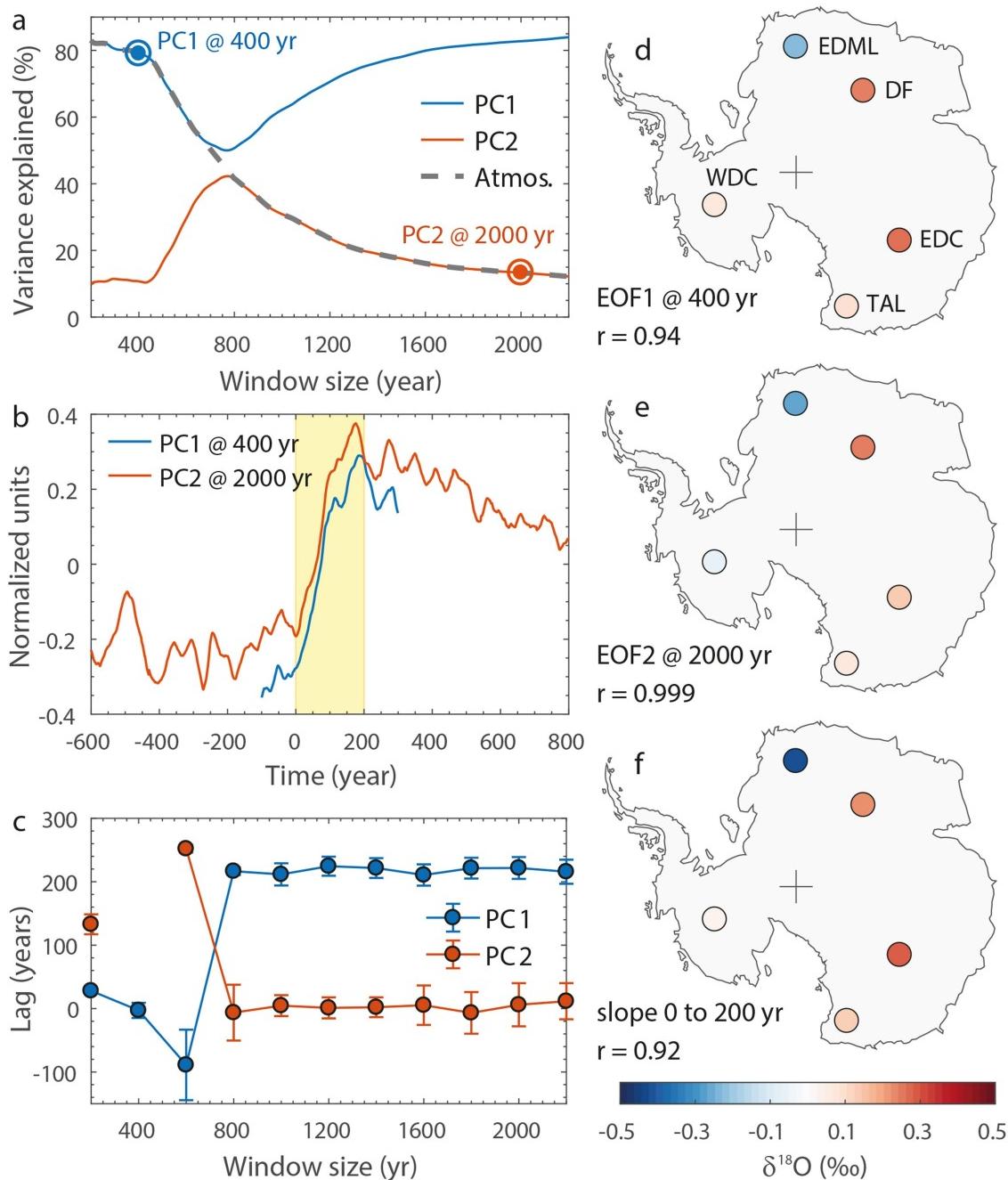
Extended Data Fig. 6 | Robustness of the atmospherically forced warming pattern. a, PCA as a function of window length, with the fraction of variance explained by PC1 and PC2. **b**, Comparison of PC1 at a 400-yr window length to PC2 at a 2,000-yr window length, showing the crossover of the atmospheric response (that is, from PC1 to PC2) as a function of window length. **c**, Lag time of the Antarctic PC1 and PC2 response as a function of window length, assessed using the BREAKFIT[55]

and RAMPFIT[56] routines (see Methods for the selection of routine). **d**, EOF1 at a 400-yr window length, expressed in units of $\delta^{18}O$ (‰). **e**, EOF2 at a 2,000-yr window length, expressed in units of $\delta^{18}O$ (‰). **f**, Slope of linear fit to $\delta^{18}O$ stacks in the interval $t = 0$ to $t = +200$ yr, shown as the change (in ‰) during these 200 years. Isotope ratios are on the VSMOW scale.
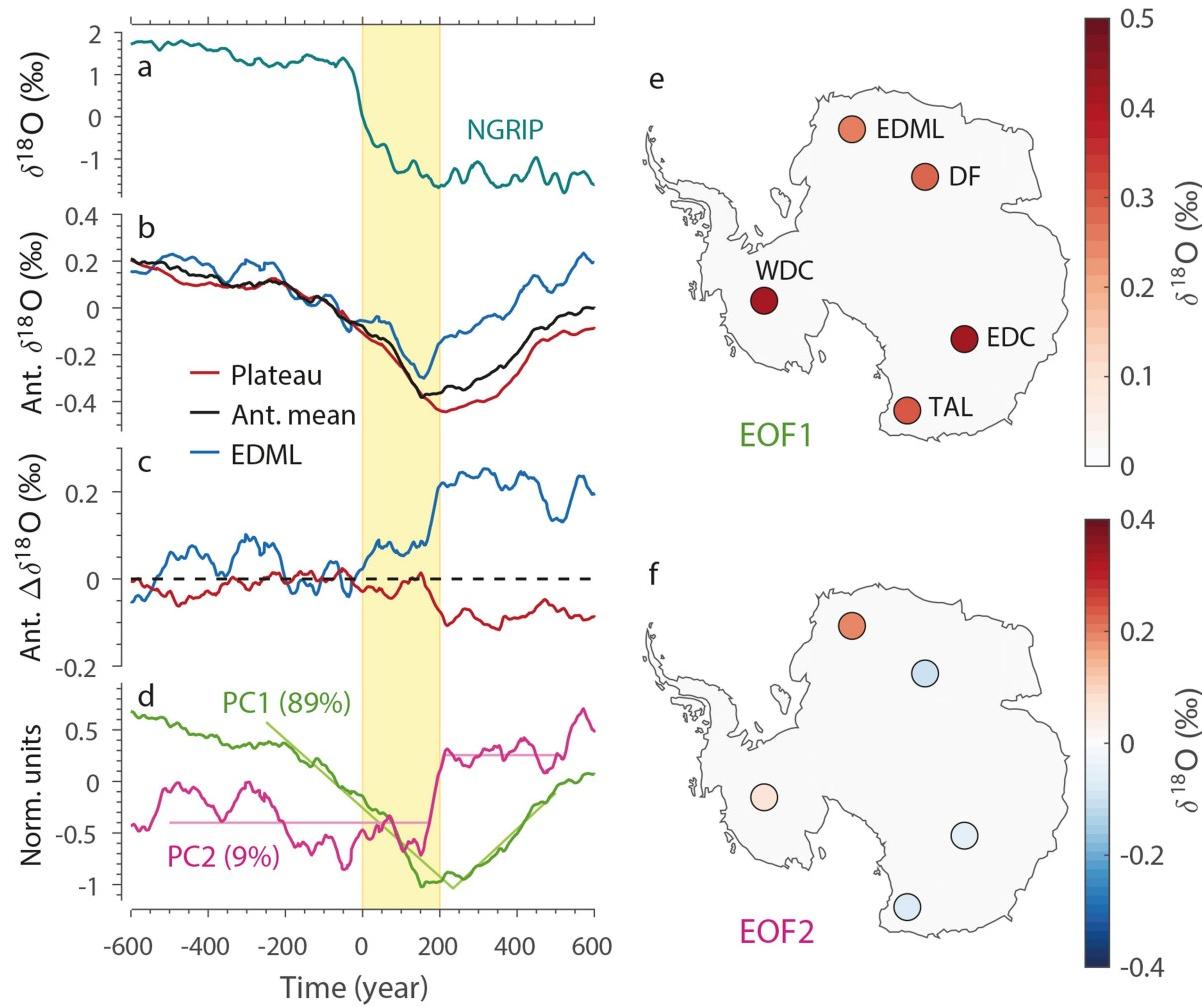
Extended Data Fig. 7 | Antarctic climate response to DO cooling.
a, Stack of NGRIP δ^{18}O. **b**, Stack of Antarctic δ^{18}O at indicated locations. **c**, As in **b**, but with the Antarctic mean subtracted. **d**, First two principal components of the Antarctic δ^{18}O stacks, with the fraction of variance explained (offset for clarity). The lines show the BREAKFIT (PC1) and RAMPFIT (PC2) fits. **e**, **f**, Empirical orthogonal functions EOF1 and EOF2 associated with PC1 and PC2 in **d**, scaled to show the magnitude in units of δ^{18}O (‰). Isotope ratios are on the VSMOW scale.
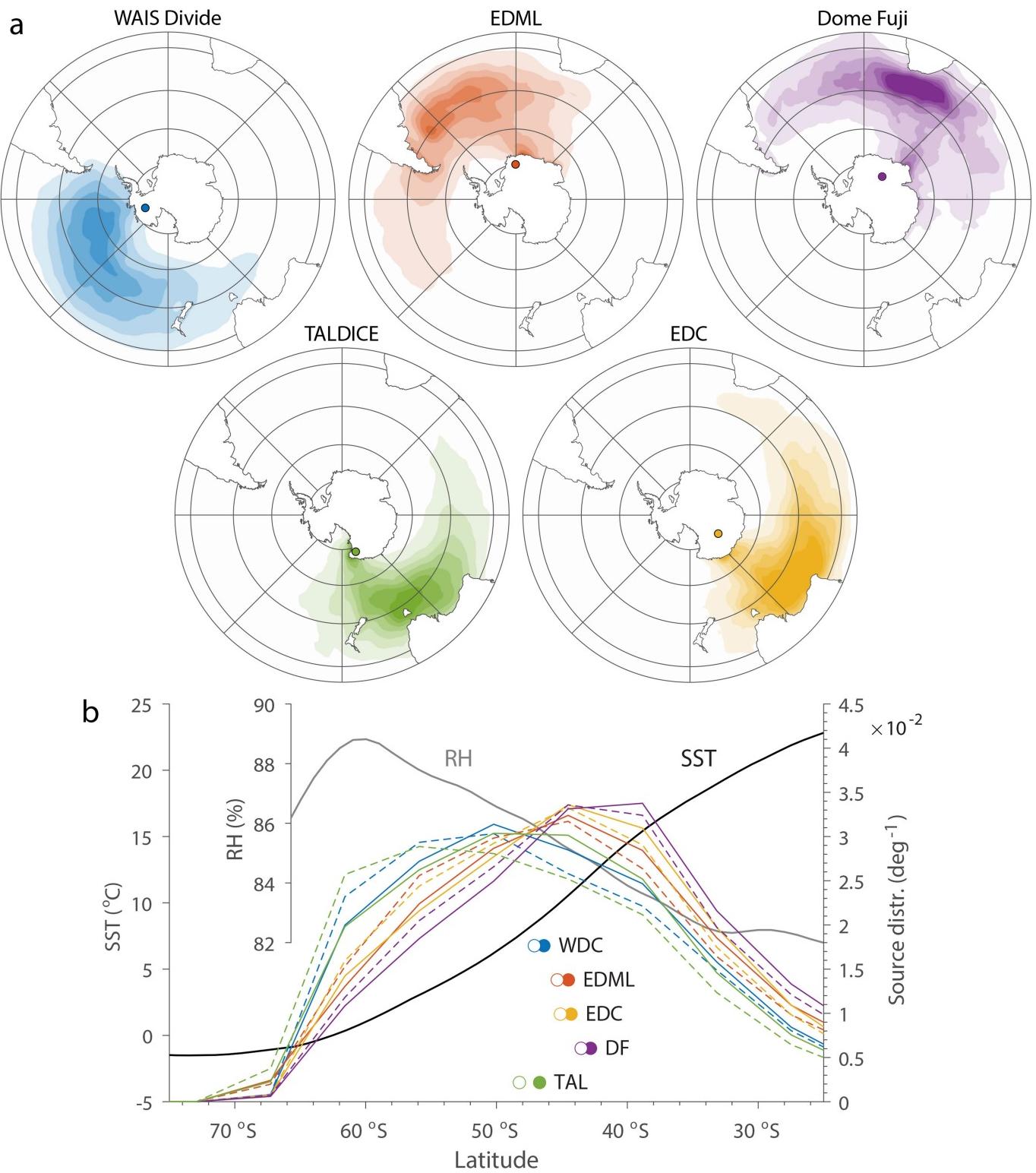
Extended Data Fig. 8 | Moisture sources of Antarctic ice cores and the SAM. a, Mass-weighted probability distribution functions of Antarctic moisture sources for the five ice cores of interest (5×10^{-5} deg^{-2} contour lines; the area-integrated probability distribution function equals 1). Distributions are calculated from reanalysis data[30,67] using a Lagrangian source diagnostic[21,66] (Methods). Parallels are plotted in 15° increments of latitude and meridians in 45° increments of longitude. **b**, SST (black)[69] and relative humidity (RH; grey)[70] as a function of latitude. The coloured curves give the latitudinal source distribution during a negative SAM phase (solid curves; SAM index <0) and a positive SAM phase (dashed curves; SAM index >0). The solid and open dots show the first moment of the source distribution during negative and positive SAM phases, respectively. We note that during a positive SAM phase, moisture sources for all core locations are located closer to the Antarctic continent. Source distribution data were obtained using water tagging experiments[8] in the CAM (Methods).
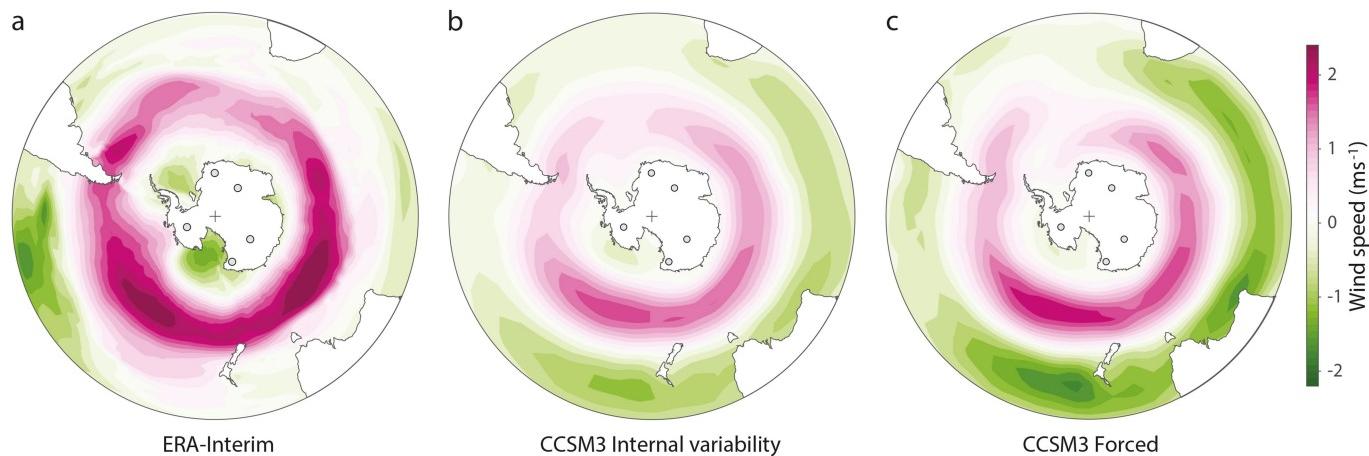
Extended Data Fig. 9 | SAM-like variability in zonal near-surface winds.
a, ERA-Interim reanalysis (1979–2016 annual means) zonal wind speed at a height of 10 m, regressed onto the SAM index (here the first principal component of sea-level pressure variability south of 20° S), expressed in metres per second per standard deviation in the index. **b**, As in **a**, but for internal variability in the CCSM3 TraCE model simulation[25,62] during glacial climate before freshwater forcing of Heinrich stadial 1 (19.5 kyr BP to 19.01 kyr BP, decadal means). **c**, As in **a**, but for the response to North Atlantic freshwater forcing in the CCSM3 TraCE model (19.1 kyr BP to 18.9 kyr BP, decadal means, with the freshwater forcing applied at 19 kyr BP).

Extended Data Table 1 | Change-point analysis of Antarctic response

Parameter	Response to:	t_{OCE}	t_{ATM1}	t_{ATM2}	Routine:
$\delta^{18}O$ stack	DO warming	226 ± 23			BREAKFIT
$\delta^{18}O$ PC1	DO warming	222 ± 24			BREAKFIT
$\delta^{18}O$ ocean	DO warming	211 ± 12			BREAKFIT
$\delta^{18}O$ stack	DO cooling	225 ± 19			BREAKFIT
$\delta^{18}O$ PC1	DO cooling	234 ± 21			BREAKFIT
$\delta^{18}O$ PC2	DO warming		28 ± 3	137 ± 3	RAMPFIT
$\delta^{18}O$ atmos	DO warming		28 ± 3	137 ± 3	RAMPFIT
d_{ln} stack	DO warming		-11 ± 34	193 ± 27	RAMPFIT
d_{ln} stack	DO warming		28 ± 14	248 ± 14	BREAKFIT (2×)
CCSM3 SAM	DO warming		13 ± 38	68 ± 48	RAMPFIT
$\delta^{18}O$ PC2	DO cooling		171 ± 5	209 ± 5	RAMPFIT
d_{ln} stack	DO cooling		6 ± 16	151 ± 19	RAMPFIT
d_{ln} stack	DO cooling		12 ± 5	166 ± 6	BREAKFIT (2×)
CCSM3 SAM	DO cooling		20 ± 28	81 ± 29	RAMPFIT

Change points are found using the BREAKFIT[55] or RAMPFIT[56] routine, as indicated. The parameter t_{OCE} is the single change point of the oceanic mode; t_{ATM1} and t_{ATM2} are the two change points of the atmospheric mode, representing the onset and ending of the shift, respectively (Methods). Stated uncertainties give the 1σ value in the fitting routine only, found using a Monte Carlo moving-block bootstrap analysis with 1,000 iterations[55,56]; the full uncertainty in the combined interpolar CH_4 synchronization and stacking procedure[15] is estimated to be around 40 yr (1σ), which can mostly be attributed to uncertainty in the WAIS Divide ice age−gas age difference, Δage[37]. The $\delta^{18}O$ PC2 and $\delta^{18}O$ atmospheric modes are identical.

LETTER

Similar cranial trauma prevalence among Neanderthals and Upper Palaeolithic modern humans

Judith Beier[1], Nils Anthes[2], Joachim Wahl[1,3] & Katerina Harvati[1,4]*

Neanderthals are commonly depicted as leading dangerous lives and permanently struggling for survival. This view largely relies on the high incidences of trauma that have been reported[1,2] and have variously been attributed to violent social behaviour[3,4], highly mobile hunter-gatherer lifestyles[2] or attacks by carnivores[5]. The described Neanderthal pattern of predominantly cranial injuries is further thought to reflect violent encounters with large prey mammals, resulting from the use of close-range hunting weapons[1]. These interpretations directly shape our understanding of Neanderthal lifestyles, health and hunting abilities, yet mainly rest on descriptive, case-based evidence. Quantitative, population-level studies of traumatic injuries are rare. Here we reassess the hypothesis of higher cranial trauma prevalence among Neanderthals using a population-level approach—accounting for preservation bias and other contextual data—and an exhaustive fossil database. We show that Neanderthals and early Upper Palaeolithic anatomically modern humans exhibit similar overall incidences of cranial trauma, which are higher for males in both taxa, consistent with patterns shown by later populations of modern humans. Beyond these similarities, we observed species-specific, age-related variation in trauma prevalence, suggesting that there were differences in the timing of injuries during life or that there was a differential mortality risk of trauma survivors in the two groups. Finally, our results highlight the importance of preservation bias in studies of trauma prevalence.

Neanderthals are commonly depicted as robust hominins who led stressful, dangerous lives[1,6–9]. Traumatic injuries, considered to be common among remains of adult Neanderthals[1], are a major piece of evidence supporting this hypothesis: not only are Neanderthals proposed to suffer from a high prevalence of trauma[2,3,10,11], but they are also thought to exhibit more traumatic injuries than early modern humans[9,12,13]. Explanations for this include violent social behaviour[3,4], a highly mobile hunter-gatherer lifestyle in glacial environments[2] and attacks by carnivores[5]. Moreover, Neanderthals are thought to show unusually high levels of head and neck injuries, attributed to their hypothesized reliance on close-range hunting leading to confrontations with large prey mammals[1]. These interpretations have important implications for reconstructions of Neanderthal palaeobiology and behaviour, and have shaped the prevailing perception of the species. However, they are largely based on anecdotal evidence, because trauma among Palaeolithic humans is often reported on a descriptive, case-by-case basis. The few systematic, quantitative studies that have been conducted to date have yielded contradictory results[2,4,11,14,15], but question the prevailing view of 'the highly traumatized Neandertal'[15].

Current research into Palaeolithic trauma suffers from several limitations. Most previous work assessed the proportional distribution of lesions throughout the body in injured Neanderthal skeletons, comparing the derived ratios to those of recent humans[1,5,15–17]. Such approaches provide insights into individual life histories, but—because they focus exclusively on the injured—cannot elucidate population-level trauma prevalence. The latter requires an examination of both injured and

uninjured individuals. Furthermore, contextual factors such as age at death, sex and skeletal preservation are rarely accounted for in these approaches[15]. These variables can markedly affect trauma prevalence variation[18–21] and lesion visibility in the fossil record, and should thus be integral to population-level analyses. Moreover, Neanderthals are routinely compared to recent humans—clinical[1] or forensic[5] samples, rodeo riders[1] and Holocene hunter-gatherers or nomads[2,4,15,16]—but only rarely to Upper Palaeolithic modern humans[17]. However, the latter are the most appropriate comparative sample, sharing similar environments and comparable mobile hunter-gatherer lifestyles. Finally, small sample sizes have hampered the validity of the statistical inferences of most previous research.

Here we assess the hypothesis of higher prevalence of cranial trauma among Neanderthals relative to Upper Palaeolithic modern humans using a population-level comparison, including contextual data and using the largest fossil dataset that is currently available. We systematically compiled published information on fossil crania from the Middle and Upper Palaeolithic of Eurasia, dating to roughly 80–20 thousand years ago (Fig. 1). Cranial injuries—considered typical for Neanderthals[1]—are a particularly reliable trauma archive, because they heal with only minor bone remodelling and therefore leave visible lesions even after full recovery[22].

For each specimen, we recorded whether trauma was present (0 or 1), the taxon (Neanderthals or Upper Palaeolithic modern humans), sex (male, female or unknown), age at death (juvenile to young adult, old adult or indeterminate), preserved skeletal element(s) (14 major cranial bones), the preservation percentage of each skeletal element ($\leq$25%, 25–50%, 50–75% and 75–100%) and location (five geographical regions within Eurasia; see Supplementary Tables 1, 2). We then used generalized linear mixed models (GLMMs) to assess differences in trauma prevalence with taxon, sex, age and preservation as explanatory variables, while accounting for variation among skeletal elements and locations.

Our systematic literature survey revealed 21 specimens with one or several cranial lesions (9 Neanderthals and 12 Upper Palaeolithic modern humans; Supplementary Table 3) in our sample of 114 specimens of Neanderthals and 90 specimens of Upper Palaeolithic modern humans (Supplementary Tables 1, 2). At the level of skeletal elements, this corresponds to 14 out of 295 cranial elements of Neanderthals, and 25 out of 541 cranial elements of Upper Palaeolithic modern humans, exhibiting at least one traumatic lesion.

We calculated separate models to predict trauma prevalence at the specimen and the skeletal-element level. Our analysis comprised two sets of four GLMM models each that were based on hierarchically nested subsets of the raw data. The first set (models 1–4; Fig. 2) followed an element-based approach, with skeletal elements being the unit of analysis; the second set (models 5–8; Fig. 3) was based on individuals (see Methods). Trauma was modelled as a binary response variable in all models, either per skeletal element or per specimen. The random component of the GLMMs comprised skeletal element and location in models 1–4, and only location in models 5–8.

[1]Paleoanthropology, Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany. [2]Animal Evolutionary Ecology Group, Institute of Evolution and Ecology, University of Tübingen, Tübingen, Germany. [3]State Office for Cultural Heritage Management Baden-Württemberg, Osteology, Konstanz, Germany. [4]DFG Center for Advanced Studies 'Words, Bones, Genes, Tools', University of Tübingen, Tübingen, Germany. *e-mail: katerina.harvati@ifu.uni-tuebingen.de

Fig. 1 | Neanderthal and Upper Palaeolithic modern human sites.
Neanderthal sites, blue triangles; Upper Palaeolithic modern human sites, red dots. Numbers in brackets indicate number of specimens/number of skeletal elements, respectively. Sites Chagyrskaya (34) and Pokrovka (74) were projected 2,670 and 2,975 km west for better visualization. The map was generated using the QGIS Geographic Information System (https://www.qgis.org) and Natural Earth (http://naturalearthdata.com/).

1 Gibraltar (1,10)	14 La Ferrassie (2,14)	27 Sarstedt (2,2)	40 Caldeirão (1,1)
2 Horá (1,1)	15 La Chapelle (1,12)	28 Kůlna (2,2)	41 Parpallo (1,11)
3 Palomas (3,9)	16 Combe Grenal (6,6)	29 Ochoz (1,2)	42 Isturitz (1,2)
4 Cova Foradà (1,3)	17 Monsempron (2,2)	30 Vindija (26,33)	43 Brassempouy (1,2)
5 Cova Negra (2,2)	18 Hortus (2,3)	31 Subalyuk (1,3)	44 Vilhonneur (1,4)
6 Gegant (1,2)	19 Fate (2,2)	32 Grotta Breuil (1,1)	45 Fontéchevade (1,1)
7 Petit-Puymoyen (3,4)	20 Ciota Ciara (1,1)	33 Guattari (3,14)	46 Cussac (1,9)
8 La Quina-Amont (9,21)	21 Cotencher (1,2)	34 Chagyrskaya (1,1)	47 Cro Magnon (3,29)
9 Pradelles/Marillac (16,19)	22 Genay (1,5)	35 Zaskalnaya VI (1,1)	48 Abri Pataud (1,14)
10 St. Césaire (1,8)	23 Spy (2,16)	36 Sakajia (1,1)	49 Mollet (1,5)
11 Fontéchevade (1,3)	24 Zeeland Ridges (1,1)	37 Shanidar (6,49)	50 Crouzade (2,3)
12 Régourdou (1,2)	25 Neanderthal (1,8)	38 Amud (3,14)	51 La Balauzière (2,8)
13 Le Moustier (1,12)	26 Warendorf (1,1)	39 Kebara (2,3)	52 Baousso da Torre (1,4)

53 Barma Grande (3,28)	66 Tapolca (1,1)	
54 Caviglione (1,13)	67 Cioclovina (1,8)	
55 Grotte des Enfants (2,28)	68 Oase (2,14)	
56 Arene Candide (1,13)	69 Muierii (2,12)	
57 Grotta Paglicci (9,20)	70 Bacho Kiro (1,1)	
58 Ostuni (1,14)	71 Buran Kaya III (3,4)	
59 Mladeč (9,42)	72 Kostenki (3,24)	
60 Predmost (11,85)	73 Sunghir (3,26)	
61 Brno (2,11)	74 Pokrovka (1,1)	
62 Pavlov (3,13)	75 Ohalo II (1,14)	
63 Dolní Věstonice (6,67)	76 el-Wad (3,3)	
64 Willendorf (1,2)		
65 Vindija (3,4)		

Model 1 comprised the full dataset of all skeletal elements ($n = 836$) to exclusively assess overall taxon differences in trauma prevalence, while ignoring the incompletely scored contextual variables. Model 2 ($n = 604$) excluded skeletal elements of unknown sex and indeterminate age, thus assessing the additional influence of age, sex, element preservation, and the interaction between age and taxon. Given trauma
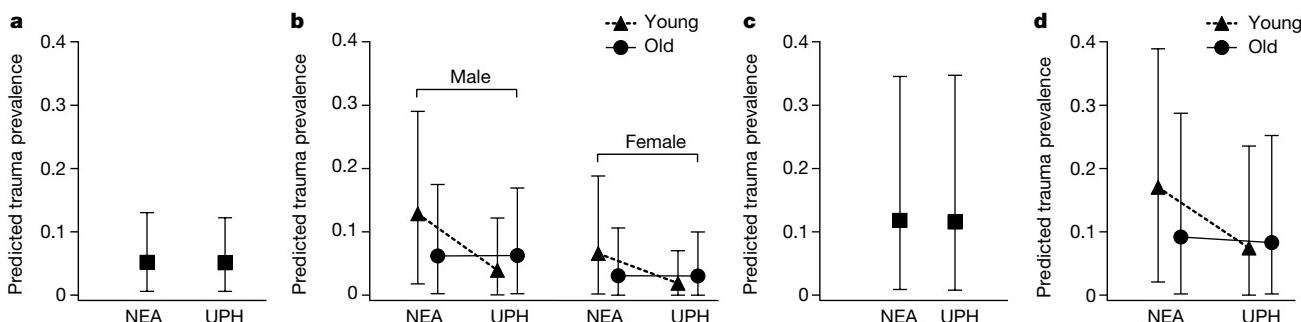
Fig. 2 | Predicted cranial trauma prevalence in skeletal elements from Neanderthals and Upper Palaeolithic modern humans. a, Model 1 includes taxon as the predictor variable (full dataset, $n = 836$). **b**, Model 2 includes the variables taxon, sex, age, element preservation and the interaction between age and taxon, but excludes sex unknown and age indeterminate skeletal elements ($n = 604$). **c**, Model 3 includes taxon as the variable, but excludes female and sex unknown skeletal elements ($n = 462$). **d**, Model 4 includes the variables taxon, age, element preservation, and the interaction between age and taxon, but excludes female, sex unknown and age indeterminate skeletal elements ($n = 407$). Predictions are given for skeletal elements when 50–75% complete; predictions for other preservation categories scale linearly. Predictions are based on posterior estimates of the four GLMMs using a Markov chain Monte Carlo (MCMC) algorithm. Sample sizes represent single skeletal elements, treated as biologically independent samples in models 1–4 (see Methods). Markers denote the predicted means, bars show lower and upper 95% credible intervals. NEA, Neanderthals; UPH, Upper Palaeolithic modern humans.
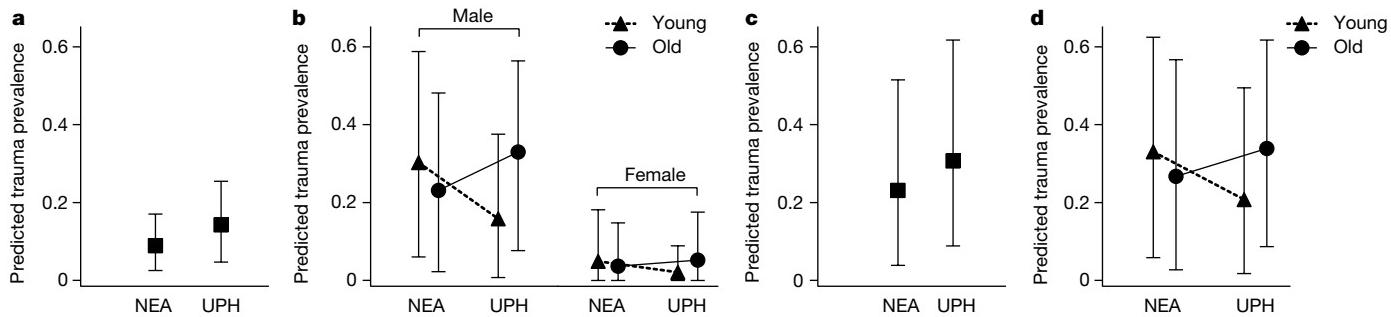
Fig. 3 | Predicted cranial trauma prevalence in individual cranial specimens from Neanderthal and Upper Palaeolithic modern humans.
a, Model 5 includes taxon as the predictor variable (full dataset, $n = 204$).
b, Model 6 includes the variables taxon, sex, age, specimen preservation and the interaction between age and taxon, but excludes sex unknown and age indeterminate specimens ($n = 89$). **c**, Model 7 includes taxon as the variable, but excludes female and sex unknown specimens ($n = 76$).
d, Model 8 includes the variables taxon, age, specimen preservation and

the interaction between age and taxon, but excludes female, sex unknown and age indeterminate specimens ($n = 59$). Predictions are given for mean specimen-preservation scores; predictions for other preservation scores scale linearly. Predictions are based on posterior estimates of the four GLMMs using a Markov chain Monte Carlo algorithm. Samples sizes in models 5–8 represent cranial specimens, comprising one or several skeletal elements of the same cranium (see Methods). Markers denote the predicted means, bars indicate lower and upper 95% credible intervals.

predominance in males, we repeated these models on male-only subsets in models 3 ($n = 462$) and 4 ($n = 407$).

Model 5 comprised all specimens ($n = 204$) and, corresponding to model 1, assessed overall taxon differences in trauma prevalence. Model 6 ($n = 89$) excluded specimens of unknown sex and indeterminate age to assess how age, sex, specimen preservation and the interaction between age and taxon affected trauma prevalence. We repeated these models for male-only subsets in models 7 ($n = 76$) and 8 ($n = 59$).

None of the models showed a quantitative difference in cranial trauma prevalence between Neanderthals and Upper Palaeolithic modern humans (taxon effect in models 1–8 in Table 1 and Figs. 2a–d, 3a–d).

Instead, we found a significantly higher prevalence of trauma in males compared to females (sex effect in models 2 and 6; Table 1 and Figs. 2b, 3b). Furthermore, trauma prevalence significantly increased with preservation status, indicating a greater probability to detect trauma on more complete skeletal elements or individuals (preservation effect in models 2, 4, 6 and 8; Table 1 and Extended Data Fig. 1a). Finally, in the element-based models, trauma prevalence varied between age classes with distinct patterns for the two taxa (age-by-taxon interaction in models 2 and 4; Table 1, Fig. 2b, d and Extended Data Fig. 1b): Neanderthals had a significantly higher prevalence of trauma when young, whereas Upper Palaeolithic modern humans showed a similar

Table 1 | Summary statistics of the GLMMs

Model	n	Predictor variable	Parameter estimates			
			Posterior mean	Lower 95% CI	Upper 95% CI	P_{MCMC}
Model 1	836[a]	Taxon	0.020	−0.889	0.933	0.965
Model 2	604[b]	Taxon	−0.060	−2.017	1.687	0.949
		Sex	1.515	0.178	2.921	0.017**
		Age	−0.973	−2.154	0.210	0.100
		Element preservation	0.866	0.232	1.514	0.006***
		Age × taxon	2.595	0.573	4.645	0.008***
Model 3	462[c]	Taxon	0.052	−1.167	1.329	0.940
Model 4	407[d]	Taxon	0.220	−1.934	2.439	0.863
		Age	−0.340	−1.553	1.050	0.605
		Element preservation	0.671	0.048	1.376	0.037**
		Age × taxon	2.149	0.048	4.355	0.046**
Model 5	204[a]	Taxon	−0.651	−1.719	0.472	0.231
Model 6	89[b]	Taxon	−0.715	−2.864	1.650	0.522
		Sex	3.533	0.865	6.397	0.002***
		Age	−1.490	−3.454	0.561	0.137
		Specimen preservation	0.882	0.054	1.730	0.032**
		Age × taxon	2.019	−1.190	5.030	0.196
Model 7	76[c]	Taxon	−0.743	−2.443	0.749	0.354
Model 8	59[d]	Taxon	−0.513	−2.902	1.858	0.660
		Age	−1.153	−3.333	0.736	0.255
		Specimen preservation	0.739	−0.106	1.623	0.082*
		Age × taxon	1.584	−1.762	4.621	0.320

Trauma prevalence was modelled using a MCMC algorithm in two model sets with four data subsets each: models 1–4 comprise skeletal elements, models 5–8 comprise cranial specimens. Parameter estimates are given as their posterior mean with 95% credible intervals (CI) and statistical significance (P_{MCMC}; ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$). See Methods for details.
[a]Full dataset.
[b]Exclusion of sex unknown and age indeterminate elements or specimens.
[c]Exclusion of female and sex unknown elements or specimens.
[d]Exclusion of female, sex unknown and age indeterminate elements or specimens.

prevalence of trauma across age cohorts. Although a similar pattern appeared to be present in the specimen-level models (Fig. 3b, d), the interaction failed to reach statistical significance.

The mean model-predicted prevalence of trauma for skeletal elements in preservation category 50–75% was between 0.03 and 0.17 (95% credible interval, 0.0002–0.39) for Neanderthals, and between 0.02 and 0.12 (95% credible interval, 0.00006–0.35) for Upper Palaeolithic modern humans (Fig. 2a–d). For specimens, predictions were calculated for the mean specimen preservation score (a proxy for skull completeness and average preservation category of its constituent elements; see Methods). These model-predicted trauma prevalence values ranged between 0.04 and 0.33 (95% credible interval, 0.000002–0.62) for Neanderthals and between 0.02 and 0.34 (95% credible interval, 0.000001–0.62) for Upper Palaeolithic modern humans (Fig. 3a–d).

On the basis of our results, we reject the hypothesis that Neanderthals exhibit more cranial trauma than Upper Palaeolithic modern humans in western Eurasia—rather, we show that the two taxa exhibited a similar overall prevalence of cranial injuries. Previously suggested values of 30–40% cranial trauma prevalence for Neanderthals[3,10] represent the very limit of the predictions of our models for Neanderthals (mean prevalence of 3–17% for skeletal elements and 4–33% for individual specimens); these values are comparable to those found for Upper Palaeolithic modern humans (2–12% for skeletal elements and 2–34% for individual specimens) and that have been reported for Mesolithic hunter-gatherers[23], Neolithic agriculturalists[24,25] and recent hunter-gatherers[26]. Nevertheless, trauma prevalence derived from skeletal remains must not be equated to the actual numbers of injuries that were experienced during an individual's lifetime and comparisons of crude trauma frequencies should be considered with caution, because the methods used for their estimation are not always comparable among studies.

The significant relationship between trauma prevalence and sex in both taxa is consistent with observations of greater trauma prevalence among males in later periods[18,21,24–27], generally explained by sex-specific differences in activities and behaviours (division of labour, initiation rites or violent conflict)[18,20,21]. Trauma prevalence was further affected by the preservation state of skeletal remains; more complete crania or cranial elements were more likely to have preserved traumatic lesions. We therefore caution against quantitative trauma analyses that do not address preservation bias.

Both taxa showed mostly healed traumata and we did not find a markedly higher prevalence of trauma among 'old' skeletal elements in either group. This finding contradicts the expectation that healed traumatic injuries accumulate with increasing age as a result of longer exposure to dangerous situations[28], given that cranial defects remain visible over long-term periods owing to the limited regenerative bridging capacity of cranial bone healing[22]. However, death assemblages are likely to deviate from such an expected accumulation pattern[29,30], because injured individuals—even if they survived their injuries—had an increased risk of dying relative to individuals who were never injured[31,32]. Thus, our observed age pattern across taxa is consistent with the well-documented increased mortality risk of trauma survivors.

An interaction between age and taxon in trauma prevalence was found by our element-based analysis. For Neanderthals, this result suggests that cranial trauma was sustained early in life (before 30 years of age) and that trauma survivors were more likely to die while still 'young'—therefore accumulating in the 'young' age cohort in the fossil record. Once a trauma is healed, it is not possible to determine when it was acquired. Therefore, Upper Palaeolithic modern humans were either less likely to sustain trauma than Neanderthals when 'young'; and/or they sustained trauma in a similar frequency when 'young', but 'young' Upper Palaeolithic modern human trauma survivors had a lower mortality risk relative to 'young' Neanderthal trauma survivors. In other words, 'young' Upper Palaeolithic modern human injured individuals had a greater probability to survive into the 'old' age cohort. Possible explanations for these patterns include cultural or individual differences in injury proneness and healing, and different long-term

consequences of healed trauma, resulting from (for example) differences in injury severity or differential treatment of the injured—which did not, however, affect the overall prevalence of trauma.

Our study addresses the controversial topic of trauma prevalence in the Palaeolithic by reassessing cranial trauma data using a state-of-the-art methodological approach. It is, to our knowledge, the largest population-level investigation of Neanderthal cranial trauma to date and accounts for differential skeletal preservation and contextual explanatory variables using Upper Palaeolithic modern humans as a comparative sample. The available evidence indicates similar overall trauma prevalence in Neanderthals and Upper Palaeolithic modern humans in western Eurasia, rejecting earlier hypotheses of highly traumatized Neanderthals. Beyond this overall similarity, our observed age-dependent differences between the taxa also suggest possible differences in the likely age of trauma acquisition or in the mortality risk of trauma survivors.

Online content

1. Berger, T. D. & Trinkaus, E. Patterns of trauma among the Neandertals. *J. Archaeol. Sci.* **22**, 841–852 (1995).
2. Underdown, S. A comparative approach to understanding Neanderthal trauma. *Period. Biol.* **108**, 485–493 (2006).
3. Courville, C. B. in *Diseases in Antiquity* (eds Brothwell, D. R. & Sandison, A. T.) 606–622 (C. C. Thomas, Springfield, 1967).
4. Hutton Estabrook, V. & Frayer, D. W. in *The Routledge Handbook of the Bioarchaeology of Human Conflict* (eds Knüsel, C. J. & Smith, M. J.) 67–89 (Routledge, London, 2014).
5. Camarós, E., Cueto, M., Lorenzo, C., Villaverde, V. & Rivals, F. Large carnivore attacks on hominins during the Pleistocene: a forensic approach with a Neanderthal example. *Archaeol. Anthropol. Sci.* **8**, 635–646 (2016).
6. Trinkaus, E. Hard times among the Neanderthals. *Nat. Hist.* **87**, 58–63 (1978).
7. Trinkaus, E. Neanderthal mortality patterns. *J. Archaeol. Sci.* **22**, 121–142 (1995).
8. Pettitt, P. B. Neanderthal lifecycles: developmental and social phases in the lives of the last archaics. *World Archaeol.* **31**, 351–366 (2000).
9. Klein, R. G. *The Human Career. Human Biological and Cultural Origins* 3rd edn (Univ. Chicago Press, Chicago, 2009).
10. Kunter, M. Gewalt- und Arbeitsverletzungen in alter Zeit. Knochenfunde als Geschichtsquelle. *Spiegel der Forschung* **3**, 70–72 (1986).
11. Nakahashi, W. The effect of trauma on Neanderthal culture: a mathematical analysis. *Homo* **68**, 83–100 (2017).
12. Mcbrearty, S. & Brooks, A. S. The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J. Hum. Evol.* **39**, 453–563 (2000).
13. Trinkaus, E., Buzhilova, A. P., Mednikova, M. B. & Dobrovolskaya, M. V. (eds) in *The People of Sunghir* 269–294 (Oxford Univ. Press, Oxford, 2014).
14. Brennan, M. U. *Health and Disease in the Middle and Upper Paleolithic of southwestern France: a Bioarcheological Study.* PhD thesis, New York Univ. (1991).
15. Hutton Estabrook, V. *Sampling Biases and New Ways of Addressing the Significance of Trauma in Neandertals.* PhD thesis, Michigan Univ. (2009).
16. Hutton Estabrook, V. Is trauma at Krapina like all other Neandertal trauma? A statistical comparison of trauma patterns in Neandertal skeletal remains. *Period. Biol.* **109**, 393–400 (2007).
17. Trinkaus, E. Neandertals, early modern humans, and rodeo riders. *J. Archaeol. Sci.* **39**, 3691–3693 (2012).
18. Larsen, C. S. *Bioarcheology. Interpreting Behavior from the Human Skeleton* (Cambridge Univ. Press, Cambridge, 1997).
19. Jurmain, R. *Stories from the Skeleton. Behavioural Reconstruction in Human Osteology* (Gordon & Breach, Amsterdam, 1999).
20. Martin, D. L. & Harrod, R. P. Bioarchaeological contributions to the study of violence. *Am. J. Phys. Anthropol.* **156**, 116–145 (2015).
21. Redfern, R. C. *Injury and Trauma in Bioarchaeology. Interpreting Violence in Past Lives* (Cambridge Univ. Press, Cambridge, 2016).
22. Campillo, D. Healing of the skull bone after injury. *J. Paleopathol.* **3**, 137–149 (1991).
23. Terberger, T. in *Frühe Spuren der Gewalt* (eds Pieck, J. & Terberger, T.) 129–154 (Landesamt für Kultur und Denkmalpflege, Schwerin, 2006).
24. Fibiger, L., Ahlström, T., Bennike, P. & Schulting, R. J. Patterns of violence-related skull trauma in Neolithic Southern Scandinavia. *Am. J. Phys. Anthropol.* **150**, 190–202 (2013).
25. Jiménez-Brobeil, S. A., du Souich, P. & Al Oumaoui, I. Possible relationship of cranial traumatic injuries with violence in the south-east Iberian Peninsula from the Neolithic to the Bronze Age. *Am. J. Phys. Anthropol.* **140**, 465–475 (2009).

26. Schwitalla, A. W., Jones, T. L., Pilloud, M. A., Codding, B. F. & Wiberg, R. S. Violence among foragers: the bioarchaeological record from central California. *J. Anthropol. Archaeol.* **33**, 66–83 (2014).

27. Cohen, H. et al. Trauma to the skull. A historical perspective from the Southern Levant (4300bce–1917ce). *Int. J. Osteoarchaeol.* **24**, 722–736 (2014).

28. Glencross, B. & Sawchuk, L. The person-years construct: ageing and the prevalence of health related phenomena from skeletal samples. *Int. J. Osteoarchaeol.* **13**, 369–374 (2003).

29. Boldsen, J. L., Milner, G. R. & Weise, S. Cranial vault trauma and selective mortality in medieval to early modern Denmark. *Proc. Natl Acad. Sci. USA* **112**, 1721–1726 (2015).

30. Milner, G. R. & Boldsen, J. L. Life not death: epidemiology from skeletons. *Int. J. Paleopathol.* **17**, 26–39 (2017).

31. Eriksson, M., Brattström, O., Larsson, E. & Oldner, A. Causes of excessive late death after trauma compared with a matched control cohort. *Br. J. Surg.* **103**, 1282–1289 (2016).

32. Mitchell, R. J., Cameron, C. M. & McClure, R. Higher mortality risk among injured individuals in a population-based matched cohort study. *BMC Public Health* **17**, 150 (2017).

Author contributions J.B., J.W. and K.H. conceived the study. J.B. collected data. J.B., J.W., K.H. and N.A. developed the methods. J.B. and N.A. analysed the data. J.B., J.W., K.H. and N.A. wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0696-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0696-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to K.H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data collection. We collected data through a comprehensive literature review and aimed at gathering a full-evidence dataset comprising all currently known fossil crania with and without traumatic lesions. We focused on Eurasian Middle and Upper Palaeolithic sites that had yielded skull remains from classic Neanderthals (around 80–30 thousand years ago) and early to mid-Upper Palaeolithic modern humans (around 35–20 thousand years ago) (Fig. 1; Supplementary Tables 1, 2 provide information on the studied specimens). We excluded specimens that consisted of only dental remains and restricted our sample to adolescent and adult specimens with a minimum estimated age-at-death of 12 years[33]. For each specimen we recorded the taxon (Neanderthal or Upper Palaeolithic modern human), sex (male, female or unknown), age (young, 12–30 years; old, >30 years; or indeterminate, if there was no further estimate published), the skeletal element with its preservation status (see 'Quantification'), and whether the skeletal element was affected by trauma (binary). Because trauma prevalence may vary across geographical regions owing to differing social or environmental conditions, we furthermore recorded the location of each specimen (five geographical regions: Iberia, south, central, east, Near East). We adopted the assignments of taxon, sex, age and the diagnoses of traumatic lesions as published by the examiners of the specimens. These literature-based assignments may be influenced by observer bias or by the use of different methods. Nevertheless, we decided in favour of a full-evidence approach based on all available published data in order to keep data collection as consistent and complete as possible. Moreover, many fossil specimens are not available for original examination, precluding a single-method-based systematic assessment. We conducted an extensive literature review seeking to combine past research with the most recent results, so as to base our data on a complete synthesis of all available evidence, representing best-practice of research in the field. Notably, we expect misclassifications of traumatic lesions, age or sex to be equally likely in Neanderthals and Upper Palaeolithic modern humans, and this therefore should not introduce systematic biases into our group comparisons. Supplementary Table 3, a catalogue of specimens with described traumata, provides detailed descriptions of each lesion as published by the respective authors. A case was recorded as (possible) trauma once an author expressed confidence that a lesion represents a trauma, or considered a traumatic origin to be an alternative explanation for an observed lesion. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during analyses.

Quantification. Skeletal preservation has a direct effect on the census of trauma prevalence, because an injury is more likely to be detected on a more complete bone[34]. In chronologically older fragments, the preservation of skeletal remains commonly deteriorates and fragmentation of both single bones and associated skeletons increases. Moreover, the assignment of fragmented and commingled remains to specific individuals is often impossible or insecure. To account for differential skeletal preservation among sites and specimens, and to remove bias between geologically older Neanderthals and younger Upper Palaeolithic modern humans, we quantified the preservation status for each of the 14 major skull bones, that is, skeletal elements, separately. These are the frontal and occipital bones, as well as the left and right elements of the parietal, temporal, maxilla, mandible, zygomatic and nasal bones. Except for the zygomatic and nasal bones, we rated the completeness of skeletal elements in four preservation categories: up to 25%, 25–50%, 50–75% and 75–100%. Owing to their small size, the left and right zygomatic and the nasal bones were rated in just two categories: up to 50% and 50–100%. We performed the quantification procedure by visually judging the preserved portion of a given skeletal element in comparison to its complete equivalent using published pictures, sketches, virtual representations and verbal anatomical descriptions. Skeletal elements for which the preservation could not be quantified were excluded from the sample. In total, we collected data on 836 skeletal elements from 204 specimens. The quantification revealed a differential preservation among skeletal remains of Neanderthals and Upper Palaeolithic modern humans, with Neanderthals being biased towards incompletely preserved skeletal remains (see Extended Data Fig. 2a–e).

Statistical methods. We predicted trauma prevalence using GLMMs. To obtain robust GLMM estimates despite a large proportion of trauma absences (zeros) in our dataset, we used a Markov chain Monte Carlo (MCMC) algorithm as implemented in the MCMCglmm package[35] for R version 3.4.3[36]. Trauma presence or absence was modelled as a binary response variable with a binomial error distribution using a logit-link function.

Our statistical analysis of trauma prevalence comprised two sets of four GLMMs on subsets of the raw data. The first set (models 1–4) followed a skeletal element-level approach, whereas the second set (models 5–8) represented an individual specimen-level approach.

Element-level models (models 1–4). We entered the two-level predictors taxon (Neanderthals or Upper Palaeolithic modern humans), age (young or old, with 30 years as the cut-off) and sex (male or female), as well as the z-transformed

four-level covariate element preservation (0.25, 0.5, 0.75 and 1) as fixed predictor variables. Visual data inspection indicated a potential for variation in the taxon effect with age class but not with sex, so we added the age-by-taxon interaction to the models.

Because traumata are not equally frequent in the different cranial regions[24,27,37], we added intercepts for skeletal element as a random component for all element-level models, enabling us to derive marginal predictions for trauma prevalence beyond element identity while statistically accounting for variation in trauma prevalence between skeletal elements. Moreover, given that trauma prevalence may vary regionally, we added location as a second random intercept to the models.

We ran four separate models to assess trauma prevalence using four data subsets and different explanatory variable combinations, while maintaining the same two random components in each case. Model 1 included taxon as the only fixed predictor. The exclusion of the other, incompletely scored, contextual predictor variables enabled us to analyse the full dataset of $n = 836$ skeletal elements. Model 2 included all fixed predictors, that is, taxon, age, sex, element preservation and the age-by-taxon interaction. We excluded all sex unknown and age indeterminate skeletal elements from model 2, resulting in a reduced sample of $n = 604$. Given a prevalence of trauma in male individuals (see Fig. 2), we reproduced these two model variants using a male-restricted data subset. In model 3 ($n = 462$), we exclusively tested for taxon differences, excluding female and sex unknown skeletal elements. Model 4 ($n = 407$) included the predictors taxon, age, element preservation and the age-by-taxon interaction. We excluded female, sex unknown and age indeterminate skeletal elements from this model.

Specimen-level models (models 5–8). As a complementary conservative approach, we repeated the above analyses on the specimen level. This overcomes potential pseudo-replication of trauma incidence when lesions extend over multiple skeletal elements of the same cranium, or a single cranium exhibits several lesions, but does not take variation in trauma incidences between skeletal elements into account.

Specimen-level models 5–8 were identical to the element-based models 1–4, respectively, as described above. Cranial trauma presence or absence, however, was here scored at the level of specimens, resulting in sample sizes of $n = 204$ in model 5, $n = 89$ in model 6, $n = 76$ in model 7 and $n = 59$ in model 8. The preservation score in these models (specimen preservation) is a combined proxy of skull completeness and its average preservation category, calculated as the sum of all available element-based preservation scores divided by 14 skeletal elements. Location was added as the only random intercept in models 5–8.

As suggested for binary response variables[38], we fixed the residual before 1 and chose an inverse Gamma prior for random effects[39]. Model parameters were chosen to maximize model fit, visible with (i) an autocorrelation value[38] between posterior parameter estimates ≤ 0.05; (ii) parameter estimates reaching conversion between four independent model chains[40] as reflected in the potential scale reduction factor <1.01; and (iii) observed trauma prevalence falling within the 95% highest posterior density intervals of their respective posterior distribution. These criteria were met after 5,100,000 MCMC iterations, a burn-in of 100,000, and a thinning interval of 1,000, resulting in approximately 5,000 samples in all posterior distributions. From these posterior distributions, we derived the highest posterior density intervals (or credible intervals) for each parameter estimate and denoted them statistically significant (***$P < 0.01$, **$P < 0.05$) or statistical trend (*$P < 0.10$) when not including zero. These intervals formed the basis for statistical inference and hypothesis testing. Plots in Fig. 2 show model predictions for element-preservation category 50–75%, plots in Fig. 3 show the predicted trauma prevalence for specimens at their mean preservation score. In both cases, predictions linearly scale with the other preservation categories, generating overall slightly larger or smaller values but no change in the effect pattern for taxon, sex, age and the age-by-taxon interaction.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. The R code used to analyse the data in this study is available upon request from the corresponding author.

Data availability

Specimen-level data that support the findings of this study are provided in Supplementary Tables 1, 2. Quantification data for skeletal elements are available from the corresponding author upon reasonable request. Source Data for Figs. 2, 3 and Extended Data Figs. 1, 2 are provided in the online version of the paper.

33. Buikstra, J. E. & Ubelaker, D. H. (eds) *Standards for Data Collection from Human Skeletal Remains. Arkansas Archeological Survey Research Series No. 44* (Arkansas Archeological Survey, Fayetteville, 1994).
34. Judd, M. A. Comparison of long bone trauma recording methods. *J. Archaeol. Sci.* **29**, 1255–1265 (2002).

35. Hadfield, J. D. MCMC Methods for Multi-Response Generalized Linear Mixed Models. The MCMCglmmR Package. *J. Stat. Softw.* **33**, 1–22 (2010).
36. R Core Team. *R: A Language and Environment for Statistical Computing* http://www.R-project.org/ (R Foundation for Statistical Computing, Vienna, 2017).
37. Walker, P. L. Cranial injuries as evidence of violence in prehistoric southern California. *Am. J. Phys. Anthropol.* **80**, 313–323 (1989).
38. Hadfield, J. MCMCglmm Course Notes https://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf (2017).
39. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ. Press, Cambridge, 2006).
40. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).

a

b
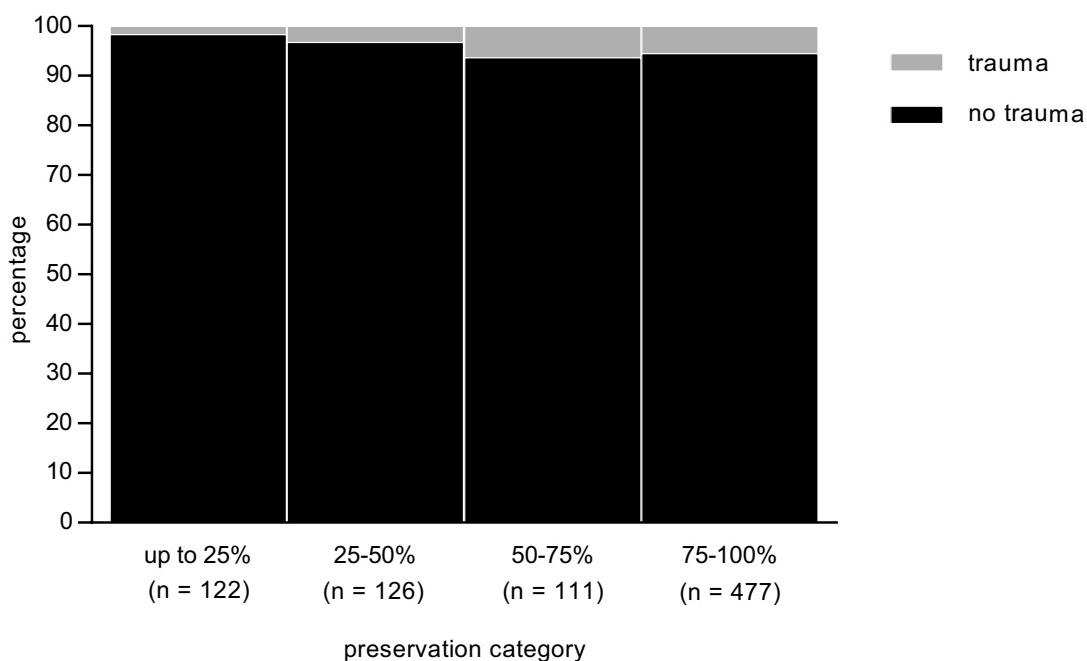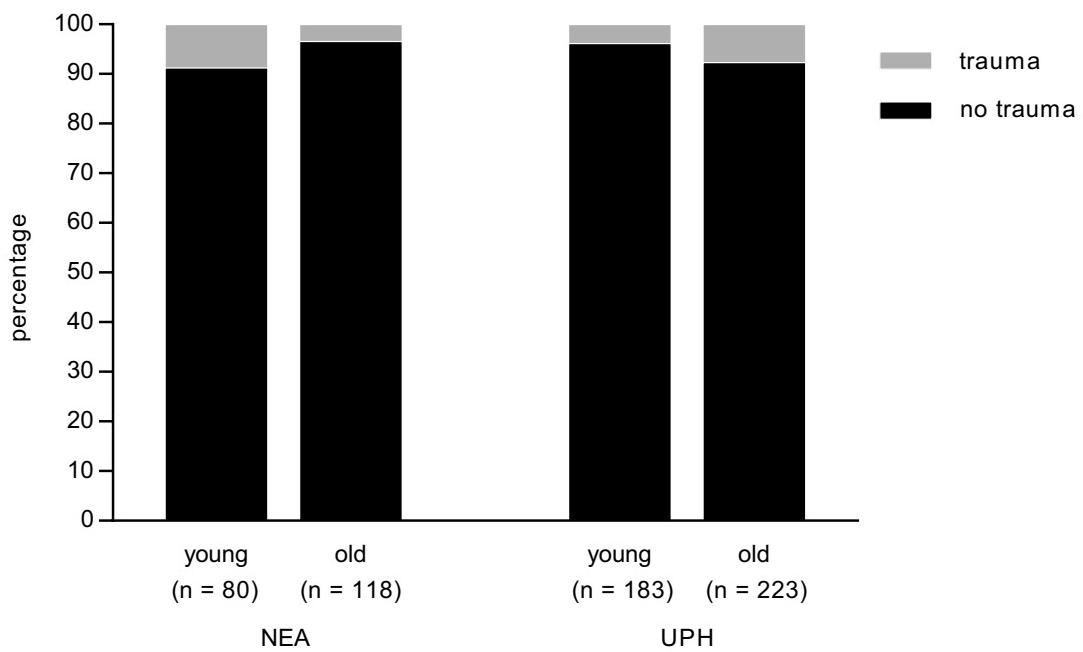
Extended Data Fig. 1 | Ratio of skeletal elements with and without trauma. a, Ratios of skeletal elements with and without trauma per preservation category for the full dataset of $n = 836$ skeletal elements. **b**, Ratios of skeletal elements with and without trauma per age cohort (young or old) and taxon (Neanderthals or Upper Palaeolithic modern humans), excluding sex unknown and age indeterminate skeletal elements ($n = 604$). Sample sizes given below bars represent numbers of skeletal elements of each subsample.
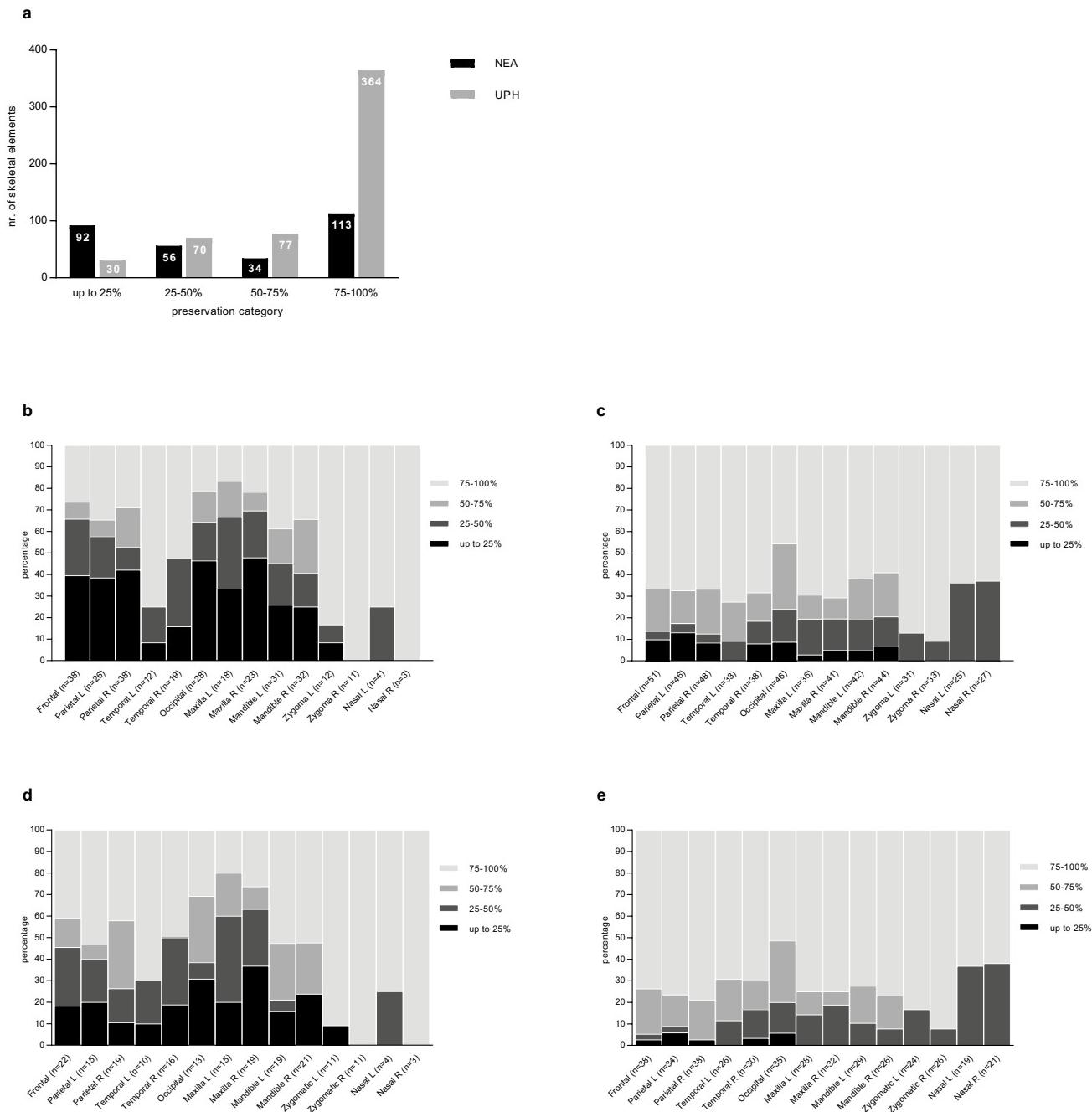
Extended Data Fig. 2 | Preservation of skeletal elements of Neanderthals and Upper Palaeolithic modern humans. a, Number of skeletal elements in each preservation category for Neanderthals and Upper Palaeolithic modern humans for the full dataset of $n = 836$ skeletal elements. b–e, Percentages of the four preservation categories for each skeletal element for Neanderthals (b; full dataset, $n = 295$ skeletal elements), Upper Palaeolithic modern humans (c; full dataset, $n = 541$ skeletal elements), Neanderthals (d; reduced dataset, excluding age indeterminate and sex unknown elements, $n = 198$) and Upper Palaeolithic modern humans (e; reduced dataset, excluding age indeterminate and sex unknown elements, $n = 406$). L and R indicate left and right, respectively.

LETTER

Trans–differentiation of outer hair cells into inner hair cells in the absence of INSM1

Teerawat Wiwatpanit[1,2,10], Sarah M. Lorenzen[1,3,10], Jorge A. Cantú[1,9,10], Chuan Zhi Foo[1,2], Ann K. Hogan[1,2], Freddie Márquez[1], John C. Clancy[1], Matthew J. Schipma[4], Mary Ann Cheatham[5,6], Anne Duggan[1,11]* & Jaime García–Añoveros[1,6,7,8,11]*

The mammalian cochlea contains two types of mechanosensory hair cell that have different and critical functions in hearing. Inner hair cells (IHCs), which have an elaborate presynaptic apparatus, signal to cochlear neurons and communicate sound information to the brain. Outer hair cells (OHCs) mechanically amplify sound-induced vibrations, providing enhanced sensitivity to sound and sharp tuning. Cochlear hair cells are solely generated during development, and hair cell death—most often of OHCs—is the most common cause of deafness. OHCs and IHCs, together with supporting cells, originate in embryos from the prosensory region of the otocyst, but how hair cells differentiate into two different types is unknown[1–3]. Here we show that *Insm1*, which encodes a zinc finger protein that is transiently expressed in nascent OHCs, consolidates their fate by preventing trans-differentiation into IHCs. In the absence of INSM1, many hair cells that are born as OHCs switch fates to become mature IHCs. To identify the genetic mechanisms by which *Insm1* operates, we compared the transcriptomes of immature IHCs and OHCs, and of OHCs with and without INSM1. In OHCs that lack INSM1, a set of genes is upregulated, most of which are normally preferentially expressed by IHCs. The homeotic cell transformation of OHCs without INSM1 into IHCs reveals a mechanism by which these neighbouring mechanosensory cells begin to differ: INSM1 represses a core set of early IHC-enriched genes in embryonic OHCs and makes them unresponsive to an IHC-inducing gradient, so that they proceed to mature as OHCs. Without INSM1, some of the OHCs in which these few IHC-enriched transcripts are upregulated trans-differentiate into IHCs, identifying candidate genes for IHC-specific differentiation.
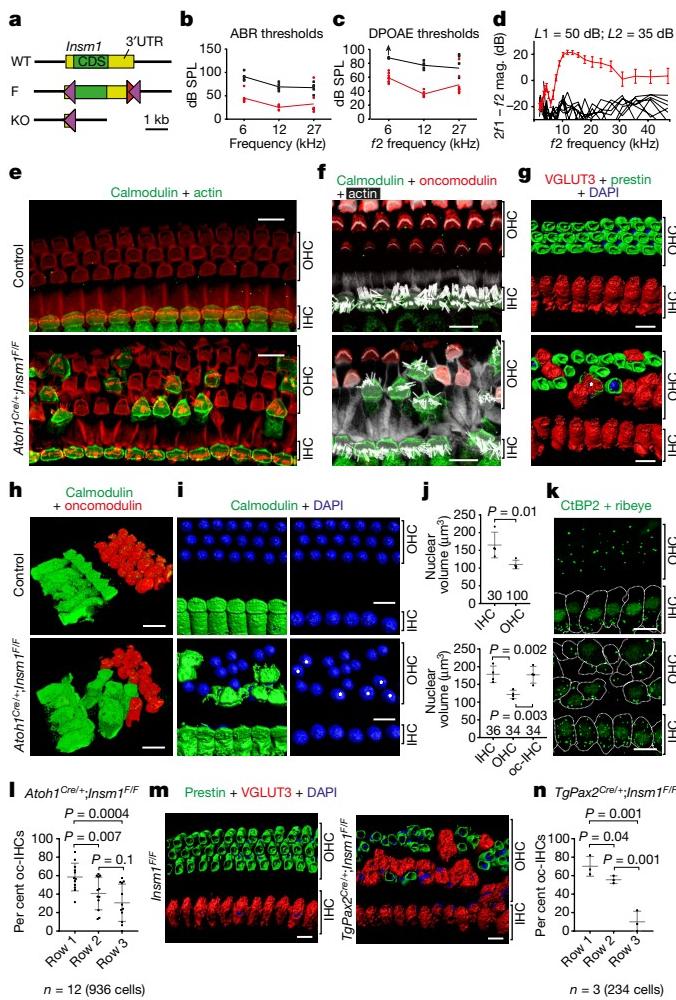
Fig. 1 | Conditional ablation of *Insm1* in hair cells results in IHC-like cells in place of OHCs. a, Wild-type (WT), floxed (F) and conditionally deleted (cKO) alleles of *Insm1*. Co-expression with Cre recombinase generates an *Insm1* knockout allele lacking its coding sequence (CDS) and 3′ untranslated region (UTR), leaving only part of the 5′UTR. Purple triangles, *LoxP* sites; red triangles, *Frt* sites. **b–d**, Hearing tests (**b, c**, mean with all values; **d**, mean ± s.e.m.). ABR thresholds (**b**), DPOAE thresholds (**c**), and iso-input functions (**d**) for the DPOAE of *Atoh1^{Cre/+}*;*Insm1^{F/F}* mice at P25–31 (black; n = 3 males and 2 females) and control littermates (red; n = 4 *Insm1^{F/F}* and 4 *Atoh1^{Cre/+}*;*Insm1^{F/+}*; 6 males and 2 females). Up arrow (**c**) indicates that maximum sound was insufficient to reach threshold. **e–l**, Immunohistochemistry in organs of Corti revealed that the *Atoh1^{Cre/+}*;*Insm1^{F/F}* mice had normal IHCs but many OHCs expressed calmodulin and not oncomodulin (**e, f, h**), had stereocilliary bundles resembling those of IHCs (f-actin labelled with phalloidin; **e, f**), expressed VGLUT3 instead of prestin (**g**; asterisk, one rare cell expressed both), had cell shapes (**h**) and large nuclei like IHCs (**i, j**, asterisks). **j**, Number of nuclei measured for each cell type are indicated under each data set. **k**, oc-IHCs had nuclear CtBP2 and a number of presynaptic ribbons (ribeye) approaching that of IHCs. **l**, In *Atoh1^{Cre/+}*;*Insm1^{F/F}* mice, oc-IHCs are more frequent in the first row of OHCs (closer to the IHCs) than the second or third rows. Data from two males, three females and seven undetermined (936 cells). **m, n**, Similar distribution of oc-IHCs in *TgPax2^{Cre/+}*;*Insm1^{F/F}* mice. n = 3 mice (234 cells). **j, l, n**, Mean ± s.e.m., one-tailed Student's *t*-tests, n = number of mice. Images and quantifications from mid-cochlear positions at P34 (**e, g, i**), P46 (**f, h**), P21, P23 and P46 (**k**), and P14 and P15 (**m**). Controls were *Insm1^{F/F}* (**e, g, i, m**) or *Atoh1^{Cre}*;*Insm1^{F/+}* (**f, j, k**) littermates. Scale bars, 10 μm. Biological replicates were used for all experiments and similar results were obtained from three or more mice per genotype.

[1]Department of Anesthesiology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA. [2]Driskill Graduate Program in Life Sciences, Northwestern University, Chicago, IL, USA. [3]NUIN Graduate Program, Northwestern University, Chicago, IL, USA. [4]Next Generation Sequencing Core, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA. [5]Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA. [6]Hugh Knowles Center for Clinical and Basic Science in Hearing and its Disorders, Northwestern University, Chicago, IL, USA. [7]Department of Neurology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA. [8]Department of Physiology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA. [9]Present address: Department of Biology, Northeastern Illinois University, Chicago, IL, USA. [10]These authors contributed equally: Teerawat Wiwatpanit, Sarah M. Lorenzen, Jorge A. Cantú. [11]These authors jointly supervised this work: Anne Duggan, Jaime García-Añoveros. *e-mail: a-duggan@northwestern.edu; anoveros@northwestern.edu
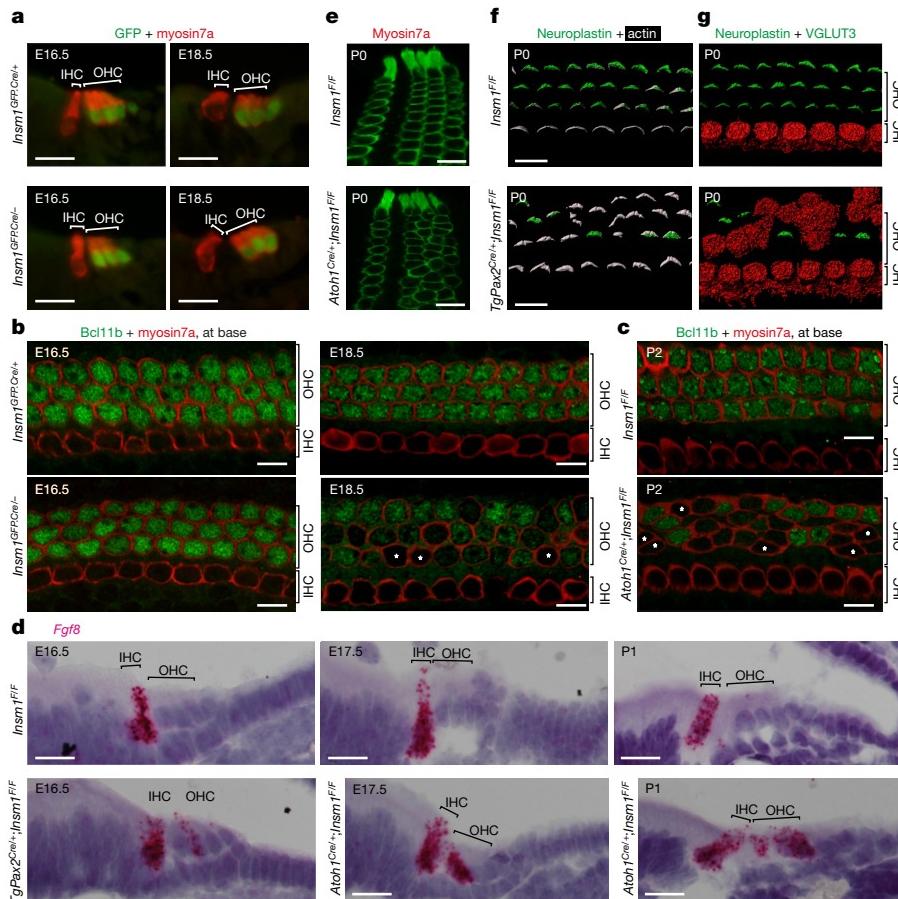
Fig. 2 | Trans-differentiation of embryonic OHCs into embryonic IHCs in the absence of INSM1. a, GFP on embryonic organs of Corti indicates that the *Insm1* promoter (expressing GFP.Cre) was active in oc-HCs (OHCs) at E16.5 and E18.5, regardless of the presence (*Insm1^GFP.Cre/+*) or absence (*Insm1^GFP.Cre/−*) of INSM1. **b, c**, Immunohistochemistry at E16.5–P2. While Bcl11b, an immature OHC-specific transcription factor, was found in all oc-HCs from control *Insm1^GFP.Cre/+* and *Insm1^GFP.Cre/−* mice at E16.5 (**b**), its expression was diminished or undetectable (asterisks) in about half of oc-HCs in mice lacking INSM1 at E18.5 (**b**) and P2 (**c**). **d**, In situ hybridization revealed that in the absence of INSM1, a subset of oc-HCs began to express *Fgf8* weakly as early as E16.5 (32.5%, 110/338 OHCs from n = 3 *TgPax2^Cre/+;Insm1^F/F* or *Atoh1^Cre/+;Insm1^F/F* mice) and strongly by E17.5–E18.5 (40%, 52/130 OHCs from n = 2

Insm1^GFP.Cre/− or *Atoh1^Cre/+;Insm1^F/F* mice) and P0–P4 (46.6%, 61/131 OHCs from n = 3 *TgPax2^Cre/+;Insm1^F/F* or *Atoh1^Cre/+;Insm1^F/F* mice). **e**, Immunohistochemistry at P0 revealed that although in *Atoh1^Cre/+;Insm1^F/F* mice there was no HC loss and OHCs retained their characteristic inclination, their cell bodies were disorganized at the nuclear level. **f**, All OHCs at the bases of cochleae from *Insm1^F/F* mice had neuroplastin in stereocilia (visualized with phalloidin). However, some OHCs at the base of cochleae from *TgPax2^Cre/+;Insm1^Flox/Flox* mice lacked neuroplastin. **g**, Conversely, the oc-HCs that lacked neuroplastin expressed VGLUT3. Images are from mid (**a–e**) or basal (**f, g**) cochlea. Scale bars, 10 μm. Biological replicates were used for all experiments and similar results obtained from three or more mice per genotype.

OHCs express *Insm1* transiently from the onset of differentiation (embryonic day 15.5 (E15.5)) to approximately postnatal day 2 (P2)[4]. Neuronal progenitors and nascent spiral ganglion neurons also express *Insm1*[4]. Because mice in which *Insm1* is completely knocked out die embryonically by E19.5[5,6], we generated an allele (*Insm1^F*) in which the entire coding sequence can be deleted (Fig. 1a, Extended Data Fig. 1). We conditionally ablated *Insm1^F* with *Atoh1^Cre*, expressed from E13.5 (three days before *Insm1*) and recombining in most cochlear hair cells and some supporting cells, but not in spiral ganglion neurons[7]. We also ablated *Insm1^F* with Tg*Pax2^Cre*, expressed earlier in the otocyst and recombining in most inner ear cells[8]. In these mice, *Insm1* was ablated before its expression in OHCs (Extended Data Fig. 2). Both *Atoh1^Cre/+;Insm1^F/F* and *TgPax2^Cre/+;Insm1^F/F* (cKO) mice displayed alterations in auditory brainstem response (ABR) thresholds that can be accounted for by shifts in distortion product otoacoustic emissions (DPOAEs), a characteristic of OHC dysfunction (Fig. 1b–d, Extended Data Fig. 3a, b). In the organs of Corti of these mice, many cells in the positions of OHCs (the outer compartment) had features of IHCs. They had large stereocilia, like IHCs, and not the shorter, W-arranged stereocillia of OHCs (Fig. 1e, f); expressed the IHC-enriched calcium buffer calmodulin and lacked OHC-specific oncomodulin (Fig. 1e, f, h,

Extended Data Fig. 3f); expressed the vesicular glutamate transporter 3 (VGLUT3), which is required for IHC presynaptic function, and lacked prestin, which is required for OHC electromotility (Fig. 1g, m); had the flask shape of IHCs rather than the cylindrical shape of OHCs; and had large nuclei, like IHCs, instead of the smaller nuclei of OHCs (Fig. 1i, j, Extended Data Fig. 3j). These nuclei harboured the transcription factor CtBP2, which is normally expressed in IHCs (Fig. 1k), and the cells contained a number of presynaptic ribbon synapses (10.6 ± 2.1 (mean ± s.d.), n = 3 mice, 39 cells) closer to that found in control IHCs (16.3 ± 0.7, n = 3 littermate controls, 30 cells), instead of the smaller number found in OHCs (1.8 ± 0.2, n = 3 mice, 90 cells) (Fig. 1k). With rare exceptions (Fig. 1g), these abnormal cells displayed all IHC features examined and lacked those of OHCs, so we termed them oc-IHCs (outer compartment IHCs).

The proportion of oc-IHCs in *Atoh1^Cre/+;Insm1^F/F* mice (42.6 ± 10.9%, n = 12 mice) and *TgPax2^Cre/+;Insm1^F/F* mice (46.0 ± 5.64%, n = 3 mice) was about half, the rest appearing as OHCs. This is not due to incomplete or delayed ablation of *Insm1*, because we did not detect *Insm1* mRNA in any OHCs of *TgPax2^Cre/+;Insm1^F/F* mice during or after the onset of expression (E16.5; Extended Data Fig. 2a, b, bottom). Notably, the oc-IHCs were more prevalent in the first hair
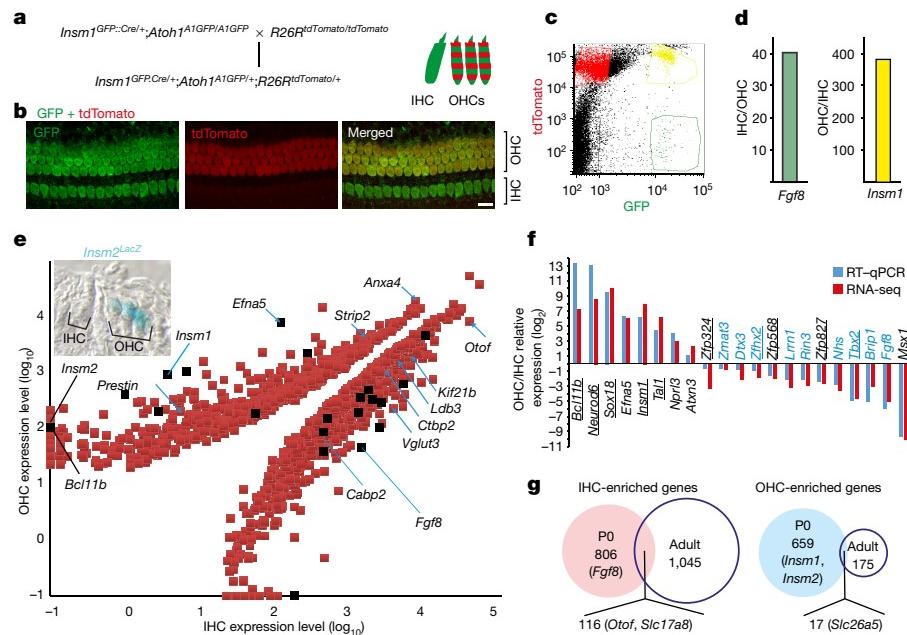
Fig. 3 | Genes preferentially expressed in immature IHCs or OHCs.
a–d, Strategy for collecting separate pools of perinatal (P0) OHCs and IHCs by FACS. **a**, Crosses for generating pups in which live IHC and OHCs can be fluorescently distinguished (green, GFP; red, tdTomato). **b**, *Insm1*[GFP.Cre/+]; *Atoh1*[A1GFP/+];*R26R*[tdTomato/+] mice express GFP in IHCs and GFP plus tdTomato in OHCs. Neurons express tdTomato but GFP from *Insm1*[GFP.Cre] has subsided. **c**, FACS separating IHCs (green) from OHCs (yellow) and neurons (red), done on six separate pools of IHCs and OHCs. **d**, RT–qPCR for IHC-specific *Fgf8* and OHC-specific *Insm1* confirm that these pools of cells are enriched for IHCs or OHCs, respectively. **e**, Logarithmic plot of genes preferentially expressed in either IHCs or OHCs based on RNA-seq values. Blue arrows indicate genes previously known to be HC subtype-specific in neonates[4,17,25–29]. We additionally confirmed perinatal OHC- or IHC-specific expression of *Insm2* with a knock-in reporter line (E17.5; top

inset). **f**, Comparison of fold difference in mRNA expression determined (for genes indicated as black boxes in **e**) by RNA-seq and RT–qPCR. Genes that encode transcription factors are underlined. IHC-specific genes potentially inhibited by INSM1 in embryonic OHCs are in blue. **g**, Venn diagrams indicating the number of genes enriched in either IHCs or OHCs of neonates versus adults (estimated from published results[30]). Representative genes are shown in parentheses. Although neonatal HCs of either type begin to show expression of some functional markers characteristic of mature cells (*Otof* and *Slc17a8* (also known as *Vglut3*) in IHCs and *Slc26a5* (also known as *Prestin*) in OHCs), the majority of hair cell type-specific genes at this early stage differ from those of the mature cells. Biological replicates were used for all experiments and similar results obtained from three or more mice per genotype.

cell row of the outer compartment than in the second or third rows (Fig. 1l, n). In principle, these oc-IHCs in mature organs of Corti lacking INSM1 could be displaced IHCs, newly generated IHCs replacing lost OHCs, IHCs born in the outer compartment, or OHCs that had trans-differentiated into IHCs. They were not displaced IHCs, as the IHC row in both cKO mice had a normal arrangement and density of IHCs (Extended Data Figs. 3c, 4a). Although during normal development cochlear hair cells are all born during embryogenesis (E12–E16)[9,10], early hair cell death can trigger the generation of hair cells from proliferating and trans-differentiating supporting cells in the first few days after birth[11–15]. This does not occur in the absence of INSM1, as hair cell density in the outer compartment (OHCs + oc-IHCs) was unaltered up to P34 (Extended Data Figs. 3d, e, 4b, c), whereas oc-IHCs were present well before that. Second, hair cells that are derived postnatally from supporting cells initially express SOX2[12–14], whereas none of the oc-IHCs of cKO pups expressed SOX2 (Extended Data Fig. 4d). Third, some postnatally produced HCs result from proliferation of supporting cells[11–14], but none of the oc-IHCs in cKO mice derived from postnatal proliferation (Extended Data Fig. 4e). These results show that oc-IHCs do not result from OHC death followed by replacement from displaced IHCs or postnatally generated hair cells. Instead, the oc-IHCs represent homeotic transformation (of mechanosensory OHCs into IHCs) due to a developmental defect in the generation or differentiation of OHCs. Either IHCs are generated in place of OHCs, or OHCs trans-differentiate into IHCs.

We examined organs of Corti from mice with conditional (*TgPax2*[Cre/+]; *Insm1*[F/F] and *Atoh1*[Cre/+];*Insm1*[F/F]) or complete (*Insm1*[GFP.Cre/−]) *Insm1* knockout during late embryogenesis, when OHCs and IHCs begin to differentiate. At E16.5, all cells in the outer compartment begin to express the earliest markers of OHCs: the *Insm1* promoter in *Insm1*[GFP.Cre/−]

embryos (which lack INSM1 but express GFP from the *Insm1* promoter[4]; Fig. 2a), and BCL11B in nuclei (Fig. 2b, c). Whereas in control mice, BCL11B expression was maintained past birth, in embryos lacking INSM1 it subsided in nearly half of outer compartment hair cells (oc-HCs) (from E18.5 to P2; Fig. 2b, c). During the same period, many oc-HCs express the early IHC marker *fgf8* (Fig. 2d). Around birth, two additional markers begin to be expressed in control mice: neuroplastin, preferentially in OHC stereocillia[16], and VGLUT3 in IHCs[17]. By comparison, in both cKO mice, many oc-HCs expressed VGLUT3 and not neuroplastin (Fig. 2f, g). Finally, although the orientation of IHCs and OHCs is maintained at birth in cKO mice, the disorganization of the OHC rows at the level of the nuclei already revealed alterations in cell shape (Fig. 2e). We conclude that in the absence of INSM1, oc-HCs are generated with early OHC features, but soon thereafter some of these cells lose these features, express early IHC markers, and proceed to differentiate into mature IHCs. This trans-differentiation of early OHCs into IHCs reveals that INSM1 is not required to initiate commitment to the OHC fate, but acts subsequently by preventing it from switching to that of IHCs. *Insm1* acts by consolidating the OHC fate, making it permanent.

Brief expression of *Insm1* is sufficient to evade phenotypic conversion (Extended Data Fig. 5). It appears that *Insm1* locks the OHC fate during a narrow developmental period. Curiously, although KO OHCs completely lack *Insm1* from their birth, fewer than half of these cells trans-differentiate into IHCs. This trans-differentiation in *TgPax2*[Cre/+];*Insm1*[F/F] and *Atoh1*[Cre/+];*Insm1*[F/F] mice is more frequent in hair cell rows closer to the IHCs than in those further away (Fig. 1l, n). This distribution reveals the existence of a gradient in the neural to abneural axis of the organ of Corti that regulates cochlear hair cell types. This gradient might induce IHC differentiation, and
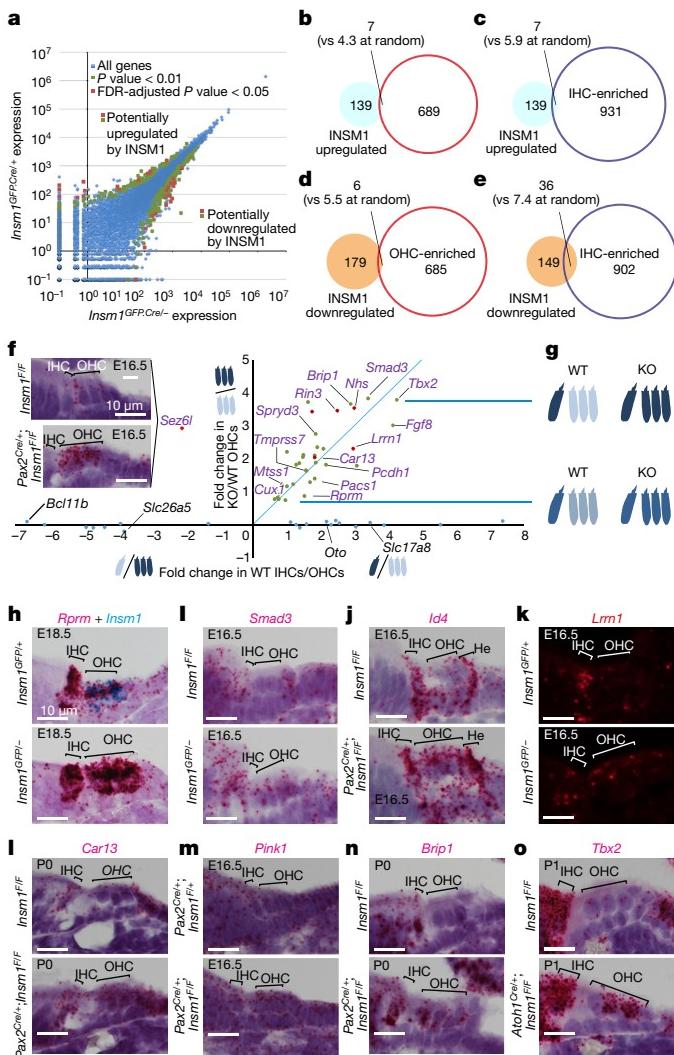
Fig. 4 | Insm1 prevents expression of a subset of immature IHC-specific genes in embryonic OHCs. a, Identification of genes potentially regulated by INSM1 in embryonic to perinatal OHCs. Plot of average expression levels determined by RNA-seq from E18.5 OHCs expressing ($Insm1^{GFP.Cre/+}$) or lacking ($Insm1^{GFP.Cre/-}$) INSM1 ($n = 3$ pools of OHCs per genotype, each from 8–12 mice). Undetected transcripts are assigned an expression of 10^{-1}. We established as cutoff either an FDR-adjusted $P < 0.05$ (red squares) or, for a less stringent selection, a raw $P < 0.01$ (green squares). Blue diamonds represent all other transcripts. **b–e**, Venn diagrams indicating overlap between IHC- or OHC-enriched genes with genes that are presumably regulated by INSM1 in OHCs. Upregulated and downregulated genes are those overexpressed in OHCs with or without INSM1, respectively. The expected number of genes that would appear by random coincidence is in parentheses. Only **e** shows a larger overlap than randomly expected, pointing to 36 IHC-specific genes that appear to be downregulated by INSM1 in OHCs. **f**, Plot of the differential expression of genes in OHCs versus IHCs and OHCs with INSM versus OHCs without INSM1. Genes are plotted as dots with colours corresponding to the P value criteria used in **a**. Expression levels in OHCs with and without INSM1 (from which KO/WT changes are estimated) are average RNA-seq values of three pools of OHCs per genotype. Expression levels in IHCs and OHCs (from which IHCs/OHCs changes are estimated) are average RNA-seq values of six pools of each cell type. Blue dots along the x-axis are examples of the many genes enriched in either IHCs or OHCs that are not affected by INSM1. Differentially expressed genes confirmed by RT–qPCR are labelled in purple. Each gene is upregulated to a similar extent in IHCs (versus OHCs) as in OHCs lacking INSM1 (versus IHCs with INSM1). **g**, Graphic interpretation. Darker shades of blue indicate higher expression. **h–o**, ISH confirms preferential expression in IHCs (and often other cells of the organ of Corti) compared with OHCs, and increased expression in OHCs lacking INSM1. Scale bars, 10 μm. Biological replicates were used for all experiments and similar results obtained from three or more mice per genotype.

INSM1 could act by preventing embryonic OHCs from responding to it.

In other developing cell types, INSM1 functions as a transcriptional activator or repressor[18–24]. We hypothesized that INSM1 directs OHCs to develop differentially from IHCs by activating OHC-specific genes or inhibiting IHC-specific genes. We first determined which genes were expressed in either differentiating hair cell type when $Insm1$ was expressed (Fig. 3), and then searched for genes regulated by INSM1 in developing OHCs (Fig. 4). For both approaches we used $Insm1^{GFP.Cre}$, in which the coding sequence of $Insm1$ is replaced by that of a fusion protein between GFP and the Cre recombinase, thereby serving as a reporter as well as a null allele[4,23]. We generated $Insm1^{GFP.Cre/+}$; $Atoh1^{A1GFP/+};R26R^{tdTomato/+}$ mice, in which all hair cells express GFP (starting at E13.5 from $Atoh1^{A1GFP}$ and, in OHCs, from $Insm1^{GFP.Cre}$) but only OHCs also express tdTomato following $Insm1^{GFP.Cre}$ expression (throughout the cochlea by E18.5[4]; Fig. 3a, b). We used these mice to sort OHCs and IHCs from neonatal (P0, approximately E19.5) organs of Corti (Fig. 3c). Using fluorescence-activated cell sorting (FACS), we collected pools of RNA from IHCs and OHCs (Fig. 3d), and then used RNA sequencing (RNA-seq) to obtain their transcriptomes (Supplementary Table 1). We thus identified 922 IHC-enriched genes and 676 OHC-enriched genes (Fig. 3e, Supplementary Tables 1–3). Among these were the 12 genes previously shown to be expressed preferentially in early IHCs or OHCs[4,17,25–29] (Fig. 3e), indicating that our approach detects most differentially expressed genes. One concern was whether genes that showed small differences in expression (twofold or less; for example, $Zmat3$), or that were detected at very low levels in one cell type only (for example, $Sox18$ and $Msx1$), were truly differentially expressed. We selected 21 transcripts (Fig. 3e), and used quantitative PCR with reverse transcription (RT–qPCR) to test for differential expression using additional pools of RNA from IHCs and OHCs. All 21 genes were confirmed to be differentially expressed, and the differences in expression were similar whether estimated by RNA-seq or RT–qPCR (Fig. 3f). We also confirmed differential expression of additional genes by methods not susceptible to potential artefacts of cell sorting and mRNA extraction: $Bcl11b$ in OHCs by immunohistochemistry (Fig. 2b); $Insm2$ in OHCs using an $Insm2^{LacZ}$ mouse line (Fig. 3e inset); and other genes by $in situ$ hybridization (ISH) as preferentially expressed in OHCs ($Neurod6$, $Sez6l$) or IHCs ($Tbx2$, $Id4$, $Rprm$, $Smad3$, $Car13$, $Brip1$, $Lrrn1$, $Pink1$) (Fig. 4g–o). These results attest to the low prevalence of false positives among the genes we estimated as being differentially expressed between immature IHCs and OHCs.

The transcriptomes of perinatal cochlear hair cells and supporting cells have been obtained, but these included a mixture of both OHCs and IHCs[25,26]. Although cell-type specific transcriptomes of mature IHCs and OHCs, obtained using microarrays, have also been published[30], we have obtained transcriptomes of these cell types before maturity, during early differentiation. A comparison of genes expressed in differentiating and mature IHCs and OHCs reveals very little overlap (Fig. 3g and Supplementary Table 4): only 5.9% of IHC-enriched and 2% of OHC-enriched genes are differentially expressed at differentiating and mature stages. These include some genes that are characteristic of the mature stage ($Vglut3$ and $Otof$ in IHCs and $Prestin$ in OHCs) but whose expression is incipient at birth. However, the vast majority of genes that were preferentially expressed in either cell type during differentiation (such as $Insm1$, $Insm2$ and $Bcl11b$ in OHCs, and $Brip1$, $Car13$ and $Fgf8$ in IHCs) are not expressed upon maturation and vice versa. Thus a complex transcriptome, involving hundreds of genes, is transiently active during OHC- and IHC-specific differentiation. It is in this genetic context that INSM1 locks the fate of OHCs so that they proceed to differentiate into mature OHCs and not IHCs.

To investigate how INSM1 drives OHC differentiation, we used RNA-seq to look for genes that were differentially expressed in differentiating OHCs with and without INSM1 ($Insm1^{GFP.Cre/+}$ versus $Insm1^{GFP.Cre/-}$) (Fig. 4a, Extended Data Fig. 6, Supplementary Table 5). We identified between 31 and 331 genes that could be differentially expressed (either upregulated or downregulated) by INSM1 (Supplementary Tables 6, 7).

Comparison of these genes with those normally enriched in OHCs or IHCs (Fig. 4b–e), combined with RT–qPCR retesting (Supplementary Table 8) and ISH (Fig. 4f–o), showed that, in OHCs, INSM1 does not activate OHC genes but rather inhibits IHC genes. No upregulated genes were confirmed by RT–qPCR and, of the 22 downregulated genes confirmed, 21 are normally preferentially expressed by IHCs. The enrichment of these genes in wild-type IHCs is similar to their upregulation in OHCs lacking INSM1 (Fig. 4f, g, Extended Data Table 1). By contrast, most genes that are differentially expressed in OHCs versus IHCs were not affected by INSM1. We conclude that INSM1 downregulates a specific subset of IHC-enriched genes in embryonic OHCs; without INSM1, those genes are expressed in embryonic OHCs, nearly half of which transdifferentiate into IHCs.

At E18.5, OHCs lacking INSM1 have not upregulated most of the early IHC-specific genes and still express early OHC-specific genes (Fig. 4f), even though many of these cells will, once differentiated, express all examined features and markers of IHCs and none of OHCs (Fig. 1e–k, m, Extended Data Fig. 3f). The small number of early IHC-specific genes (21 of 922, about 2%) that were upregulated in embryonic OHCs lacking INSM1 are likely to represent an early step in the genetic cascade that leads to their complete transformation into IHCs. As oc-HCs expressing these few genes differentiate into IHCs, these genes are likely to be required for IHC differentiation. Hence, in addition to identifying *Insm1* as a critical gene for OHC differentiation, our results also identify candidate genes for regulating the specific differentiation of IHCs. Because all OHCs express *Insm1*, but in its absence fewer than half trans-differentiate into IHCs, we expected two patterns of misexpression by ISH (Fig. 4g–o). Some genes (*Rprm*, *Id4*, *Lrrn1, Car13*, *Pink1* and *Brip1*; Fig. 4h–n) were upregulated in all OHCs lacking INSM1, as expected if they were repressed by INSM1. These must include the genes whose disinhibition in the absence of INSM1 renders embryonic OHCs susceptible to the gradient that induces IHC trans-differentiation. Other genes (*Fgf8* and *Tbx2*; Figs. 2d, 4o) were upregulated only in fewer than half of oc-HCs—presumably those that would trans-differentiate into IHCs. These genes are some of the earliest expressed in IHCs, and are likely to include regulators of IHC differentiation.

Our results reveal homeotic transformation of OHCs into IHCs in the absence of INSM1, identify the genes initially misregulated by ablation of *Insm1*, and provide a genetic mechanism for differentiation of these two cell types: nascent OHCs transiently express *Insm1*, which represses (directly or indirectly) a core set of early IHC-specific genes and renders the cells insensitive to an IHC-inducing gradient; this consolidates the fate of OHCs by preventing their trans-differentiation into IHCs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0570-8.

1. Puligilla, C. & Kelley, M. W. Building the world's best hearing aid; regulation of cell fate in the cochlea. *Curr. Opin. Genet. Dev.* **19**, 368–373 (2009).
2. Groves, A. K., Zhang, K. D. & Fekete, D. M. The genetics of hair cell development and regeneration. *Annu. Rev. Neurosci.* **36**, 361–381 (2013).
3. Basch, M. L., Brown, R. M. II, Jen, H.-I. & Groves, A. K. Where hearing starts: the development of the mammalian cochlea. *J. Anat.* **228**, 233–254 (2016).
4. Lorenzen, S. M., Duggan, A., Osipovich, A. B., Magnuson, M. A. & García-Añoveros, J. Insm1 promotes neurogenic proliferation in delaminated otic progenitors. *Mech. Dev.* **138**, 233–245 (2015).
5. Gierl, M. S., Karoulias, N., Wende, H., Strehle, M. & Birchmeier, C. The zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic beta cells and intestinal endocrine cells. *Genes Dev.* **20**, 2465–2478 (2006).
6. Rosenbaum, J. N., Duggan, A. & García-Añoveros, J. Insm1 promotes the transition of olfactory progenitors from apical and proliferative to basal, terminally dividing and neuronogenic. *Neural Dev.* **6**, 6 (2011).
7. Yang, H., Xie, X., Deng, M., Chen, X. & Gan, L. Generation and characterization of Atoh1-Cre knock-in mouse line. *Genesis* **48**, 407–413 (2010).
8. Ohyama, T. & Groves, A. K. Generation of Pax2-Cre mice by modification of a Pax2 bacterial artificial chromosome. *Genesis* **38**, 195–199 (2004).
9. Ruben, R. J. Development of the inner ear of the mouse: a radioautographic study of terminal mitoses. *Acta Otolaryngol.* **220**, 1–44 (1967).
10. Chen, P. & Segil, N. p27(Kip1) links cell proliferation to morphogenesis in the developing organ of Corti. *Development* **126**, 1581–1590 (1999).
11. Kelley, M. W., Talreja, D. R. & Corwin, J. T. Replacement of hair cells after laser microbeam irradiation in cultured organs of corti from embryonic and neonatal mice. *J. Neurosci.* **15**, 3013–3026 (1995).
12. Cox, B. C. et al. Spontaneous hair cell regeneration in the neonatal mouse cochlea in vivo. *Development* **141**, 816–829 (2014).
13. Bramhall, N. F., Shi, F., Arnold, K., Hochedlinger, K. & Edge, A. S. B. Lgr5-positive supporting cells generate new hair cells in the postnatal cochlea. *Stem Cell Reports* **2**, 311–322 (2014).
14. Hu, L., et al. Diphtheria toxin-induced cell death triggers Wnt-dependent hair cell regeneration in neonatal mice. *J. Neurosci.* **36**, 9479–9489 (2016).
15. White, P. M., Doetzlhofer, A., Lee, Y. S., Groves, A. K. & Segil, N. Mammalian cochlear supporting cells can divide and trans-differentiate into hair cells. *Nature* **441**, 984–987 (2006).
16. Mburu, P. et al. Whirlin complexes with p55 at the stereocilia tip during hair cell development. *Proc. Natl Acad. Sci. USA* **103**, 10973–10978 (2006).
17. Flores, E. N. et al. A non-canonical pathway from cochlea to brain signals tissue-damaging noise. *Curr. Biol.* **25**, 606–612 (2015).
18. Breslin, M. B., Zhu, M., Notkins, A. L. & Lan, M. S. Neuroendocrine differentiation factor, IA-1, is a transcriptional repressor and contains a specific DNA-binding domain: identification of consensus IA-1 binding sequence. *Nucleic Acids Res.* **30**, 1038–1045 (2002).
19. Jia, S., Wildner, H. & Birchmeier, C. Insm1 controls the differentiation of pulmonary neuroendocrine cells by repressing Hes1. *Dev. Biol.* **408**, 90–98 (2015).
20. Lan, M. S. & Breslin, M. B. Structure, expression, and biological function of INSM1 transcription factor in neuroendocrine differentiation. *FASEB J.* **23**, 2024–2033 (2009).
21. Liu, W.-D., Wang, H.-W., Muguira, M., Breslin, M. B. & Lan, M. S. INSM1 functions as a transcriptional repressor of the neuroD/β2 gene through the recruitment of cyclin D1 and histone deacetylases. *Biochem. J.* **397**, 169–177 (2006).
22. Wang, H.-W. et al. Identification of an INSM1-binding site in the insulin promoter: negative regulation of the insulin gene transcription. *J. Endocrinol.* **198**, 29–39 (2008).
23. Osipovich, A. B. et al. *Insm1* promotes endocrine cell differentiation by modulating the expression of a network of genes that includes *Neurog3* and *Ripply3*. *Development* **141**, 2939–2949 (2014).
24. Jia, S. et al. Insm1 cooperates with Neurod1 and Foxa2 to maintain mature pancreatic β-cell function. *EMBO J.* **34**, 1417–1433 (2015).
25. Cai, T. et al. Characterization of the transcriptome of nascent hair cells and identification of direct targets of the Atoh1 transcription factor. *J. Neurosci.* **35**, 5870–5883 (2015).
26. Scheffer, D. I., Shen, J., Corey, D. P. & Chen, Z.-Y. Gene expression by mouse inner ear hair cells during development. *J. Neurosci.* **35**, 6366–6380 (2015).
27. Yang, T. et al. Expression and localization of CaBP Ca²⁺ binding proteins in the mouse cochlea. *PLoS One* **11**, e0147495 (2016).
28. Jacques, B. E., Montcouquiol, M. E., Layman, E. M., Lewandoski, M. & Kelley, M. W. Fgf8 induces pillar cell fate and regulates cellular patterning in the mammalian cochlea. *Development* **134**, 3021–3029 (2007).
29. Pirvola, U. et al. FGFR1 is required for the development of the auditory sensory epithelium. *Neuron* **35**, 671–680 (2002).
30. Liu, H. et al. Characterization of transcriptomes of cochlear inner and outer hair cells. *J. Neurosci.* **34**, 11085–11095 (2014).

Author contributions J.G.-A. and A.D. conceived the project. T.W., S.M.L., J.A.C., C.Z.F., A.K.H., F.M., J.C.C., M.A.C., A.D. and J.G.-A. performed experiments. T.W., S.M.L., J.A.C., C.Z.F., M.J.S., M.A.C., A.D. and J.G.-A. analysed data. J.G.-A. and T.W. wrote the manuscript.

METHODS

Ethics and animals. All animal care and procedures were in strict accordance with the *Guide for the Care and Use of Laboratory Animals* published by the National Institutes of Health and were approved by Northwestern University's Institutional Animal Care and Use Committee (Animal Study Protocols IS00001281 and IS00000593). *Mus musculus* were used in this study. Animals were of CD1 and C57BL/6 genetic background. Similar numbers of male and female animals were used for all analyses, and these are stated in the manuscript. When sex was undetermined, as in the case of embryos, all the embryos were blindly selected. For embryonic samples, we used animals at E16.5, E17.5 and E18.5. For neonatal animals, the average age was 9.27 ± 5.6 (s.d.) days old. For adult animals, the average age was 34.32 ± 6.3 days old. All tests were performed in mutant and littermate control mice, which were otherwise randomly chosen from the litters. Although the phenotype is obvious, quantifications of nuclear size, ribbon synapse densities, cell densities (Fig. 1, Extended Data Figs. 3, 4) and hearing tests (Fig. 1, Extended Data Figs. 3, 5) were performed without knowledge of genotype. Transcriptomic analyses were carried out by a biostatistician from a core facility with no knowledge of genotype.

Generation of the *Insm1* floxed allele for conditional ablation. The *Insm1* targeting construct was generated using a genomic BAC clone, 439G2, from the mouse 129/SvEv genomic BAC library, RPCI-22. The *Insm1* gene, including the coding sequence, 5′ and 3′ UTRs, a 2,790-bp 5′ homologous sequence and a 4,098-bp 3′ homologous sequence, was subcloned into the pL253 vector (IA1-pL253) using recombineering, as described previously[31]. The recombined clone, IA1-pL253, was further modified using recombineering to add a *LoxP* recombination site immediately downstream of the 5′UTR but before the Kozak sequence. A second *Frt-NEO-Frt-LoxP* site was recombined immediately downstream of the 3′UTR. The completed targeting vector was sequence verified and sent to the Northwestern Transgenic and Targeted Mutagenesis Laboratory (Chicago, IL) for electroporation into SvEv 129 mouse embryonic stem cells.

Using Q5 High-Fidelity Polymerase with GC Enhancer (NEB Catalog:M0491) and the primers WT: TCTTAGATTCTGCCCTTTCTGACAG; CKO: CCAAGGAGATGACCACGCATAG; and R2: CTCTTGTAGGGCCTCCTGTG, we performed a PCR to identify recombinant clones. Conditions for Thermal-cycler were: step 1: 98 °C, 3:00 min; step 2: 98 °C, 0:10 min; step 3: 65 °C, 0:30 min, −1 °C per cycle; step 4: 72 °C, 6:45 min; repeat step 2, 10×; step 5: 98 °C, 0:10 min; step 6: 60 °C, 0:30 min; step 7: 72 °C, 6:45 min, repeat step 5, 25×; step 8: 72 °C, 10:00 min; step 9: 4 °C. Expected sizes for wild-type allele using primers WT to R2 was 6,163 bp. Expected size for recombinant clones using CKO-Reverse was 6,145 bp. We screened a total of 439 clones and identified 5 recombinants.

We further screened these five ES cell clones for recombination upstream of the 5′ *LoxP* site. DNA from selected recombinant clones was digested with the restriction enzyme SpeI (NEB Cat:R0133) and homologous recombination was confirmed by Southern hybridization. DNA was visualized using a 168-bp radio-labelled probe (as described[6]). The expected band sizes for wild-type and conditional knockout alleles are 18,162 bp and 14,333 bp, respectively. All five ES clones contained a targeted allele of *Insm1*.

These clones were used for the generation of mosaic embryos, which were implanted into surrogate mothers by the Northwestern Transgenic and Targeted Mutagenesis Laboratory (Chicago, IL). From one of these clones (B3) we first generated chimeric mice, which were mated to mice expressing the FlpE recombinase (B6-Tg(CAG-FLPe)36[32]) to delete the NEO cassette flanked by *Frt* sites and thus generate mice with a floxed allele of *Insm1*.

Hearing tests. During testing, mice of both sexes aged P25–P31 were anaesthetized with ketamine and xylazine (120 mg/kg and 10 mg/kg, respectively, intraperitoneal (IP)) and their body temperature maintained using a heating blanket. In order to assay OHC function, DPOAEs were recorded using a custom probe equipped with a sensitive microphone (Knowles Electronics, FG-3652-CX). Responses were analysed using Emission Averager (EMAV)[33]. Because the probe can be placed close to the eardrum, sound calibrations in the ear canal of each mouse were performed out to 48 kHz using a chirp stimulus generated in System Response (SysRes)[34]. All signals were generated using a CardDeluxe 24-bit sound card with a sampling rate of 96 kHz. Iso-input functions ($f2/f1 = 1.2$) at $L1 = 50$ and $L2 = 35$ dB were recorded for $f2$ frequencies between 2 and 47 kHz, thereby covering most of the mouse audiogram. Input–output functions were also acquired for various $f2$ frequencies (6, 12, 27 kHz), where $L1 = L2 + 10$ dB. Thresholds for $2f1$–$f2$ were then calculated and represent the level of $f1$ that produces a DPOAE of 0 dB. After emission testing, neural responses were measured by collecting ABRs using tone-burst stimuli. The threshold was determined by noting the level at which the ABR waveform disappeared into the noise. For these experiments, sound calibration was obtained using a real pinna coupler[35]. Further details were provided in a previous publication[36].

Tissue collection and preparation. Neonatal mice were killed by decapitation, and cochleae dissected in cold HBSS with calcium and magnesium (Gibco). For embryos, timed pregnant dams were killed by isoflurane overdose followed by cervical dislocation. Their abdomens were opened to expose the uterus, which was dissected in cold HBSS with calcium and magnesium (Gibco); embryos were then collected and their cochleae removed. After dissection, neonatal and embryonic cochleae were processed depending on future use. For immunohistochemistry, embryonic and neonatal cochleae were fixed in 4% paraformaldehyde for 2 h at room temperature. For older tissues (>P20), mice were cardiac-perfused with 4% paraformaldehyde, and cochleae were dissected and post-fixed in 4% paraformaldehyde for 2 h at room temperature. Cochleae from animals older than P5 were decalcified in 10% EDTA, pH 7.4, at 4 °C until needed. Organs of Corti were dissected out from the cochleae into one apical, two middle and one basal sections using a whole-mount surface preparation method[37]. Frozen sections were processed as described[4].

Immunohistochemistry. Whole-mount organ of Corti sections were processed for immunohistochemistry as described previously[38]. Primary antibodies were mouse anti-calmodulin (1:100, C-7055, Sigma Aldridge), goat anti-oncomodulin (1:200, sc-7446, Santa Cruz), rabbit anti-prestin (1:1,000, from J. Zheng, Northwestern University), guinea pig anti-VGLUT3 (1:2,500, from R. Edwards, University of California, San Francisco), mouse anti-CtBP2 (1:400, 612044, BD Biosciences), rabbit anti-myosin7a (1:800, 25-6790, Proteus Biosciences), sheep anti-neuroplastin (1:150, AF7818, R&D Systems), mouse anti-BCL11B (1:400, ab18465, abcam), and goat anti-SOX2 (1:500, sc-17320, Santa Cruz). For BCL11B immune-labelling on whole-mount cochlea, we performed antigen retrieval by incubating samples in 10 mM sodium citrate, pH 6 with 0.25% Triton X-100 for 20 min at 92 °C and cooling for 30 min at room temperature before blocking. For CtBP2 and SOX2 immuno-labelling, samples were prepared using a freeze–thaw method. In brief, organ of Corti sections were incubated in 30% sucrose at room temperature for 20 min, put in −80 °C for 5 min and thawed at room temperature for 20 min; sucrose was rinsed off with PBS before blocking and incubation with primary antibodies at 37 °C overnight. Nuclei were counterstained with 1:1,000 DAPI or 1:2,000 Hoechst 33342.

X-gal staining. X-gal staining to detect β-galactosidase expression on sections of *Insm2*[LacZ] embryos was performed as described[39].

Cell proliferation assay. In order to label hair and supporting cells generated from progenitors that proliferated postnatally, neonates were injected twice daily from P0 to P5 with the thymidine analogue 5-ethynyl-2′-deoxyuridine (EdU; 50 mg/kg in sterile saline). EdU incorporation into DNA was detected using Click-iT Plus EdU Alexa Fluor 555 Imaging Kit (Thermo Fisher Scientific) according to the manufacturer's manual. Following EdU detection, the samples were immune-labelled with antibodies as described above.

Image acquisition and analysis. We acquired images on either a Nikon A1 or A1R+ Confocal imaging system using a 100× objective. 3D renderings were generated using NIS Elements AR4.60.00 (Nikon) and Imaris X64 8.4.1 (BitPlane) software. Nuclei and ribbon synapses were measured using built-in analysis functions on Imaris. After acquisition, we identically processed image pairs of control and knockout samples. This included adjustment for brightness, contrast and parameters for 3D volume and surface renderings of all images.

FACS. For collection of OHCs or IHCs, organs of Corti were dissected from E18.5 embryos (*Insm1*[GFP.Cre/−] or *Insm1*[GFP.Cre/+]) or P0 neonates (*Insm1*[GFP.Cre/+]; *Atoh1*[A1GFP/+];*R26R*[tdTomato/+]) in ice-cold HBSS with calcium and magnesium (Gibco). A portion of tail from each embryo and neonate was collected for genotyping. Organs of Corti were washed three times in cold 1×PBS and then they were digested in 0.33 U/ml papain, 0.5 mM EDTA and 1 mM L-cystine in EBSS for 10 min at 37 °C, rinsed 3 times in 2% FBS and mechanically dissociated by gentle trituration (~100–150× with a P1000 pipet). Cell suspensions were kept on ice until FACS sorting on a BD FACS Aria 4 flow cytometer through a 100-μm nozzle at speed 2 (<100 events/s). Hair cell populations were collected into RLT buffer (Qiagen, Valencia, CA). RNA was then isolated from cells using Qiagen RNeasy Plus Micro Kit or cells were stored at −80 °C until RNA isolation. Isolated RNA was evaluated for quality and concentration on a BioAnalyzer and stored at −80 °C.

qRT–PCR. RT–PCR was performed using either SYBR Green or TaqMan systems. Total RNA was extracted from pools of hair cells collected from E18.5 *Insm1*[GFP.Cre/−], E18.5 *Insm1*[GFP.Cre/+], and P0 *Insm1*[GFP.Cre/+];*Atoh1*[A1GFP/+];*R26R*[tdTomato/+] mice through FACS. RNA was extracted using an RNeasy Plus Micro Kit (Qiagen) according to the manufacturer's instructions. RNA quality was determined with a BioAnalyzer through NUSeq Core Facility, Northwestern University, Chicago, IL.

For SYBR Green qRT–PCR, we used ~3,000 pg total RNA from each hair cell pool for first strand cDNA synthesis using iScript reverse transcription supermix (Bio-Rad) according to the manufacturer's manual. We then performed RT–qPCR with ~200 pg of first strand cDNA using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) in triplicate on a CFX Connect Real-Time PCR Detection System (Bio-Rad) using a 40-cycle protocol.

For TaqMan qRT–PCR, we used 1 μg total RNA from each HC pool for first strand cDNA synthesis using SuperScript VILO cDNA Synthesis Kit with ezDNase

Enzyme (Applied Biosystems) according to the manufacturer's instructions. First strand cDNA was subjected to ezDNase inactivation using 1 µl of 100 mM DTT per reaction. Prior to qRT–PCR, we performed pre-amplification of first strand cDNA using TaqMan PreAmp Master Mix (Applied Biosystems) according to the manufacturer's instructions. We then performed qRT–PCR using 1 µg diluted pre-amplified cDNA (1:20 in TE buffer) per reaction in triplicate on a QuantStudio 7 flex Real-Time system (Applied Biosystems) using a 14-cycle protocol at NUSeq Core Facility, Northwestern University, Chicago, IL.

All primers used for qRT–PCR in this study were designed and pre-mixed to their optimal concentrations by BioRad and Applied Biosystems. qRT–PCR reactions were performed according to the corresponding manufacturer's instructions.

RNA-seq and transcriptome analysis of embryonic OHCs and P0 hair cells. To purify enough RNA for deep sequencing and to analyse results statistically in order to determine the IHC and OHC transcriptomes, we collected by FACS six separate pools of IHCs (700–1,100 cells per pool) and six of OHCs (2,800–3,700 cells per pool) from $Insm1^{GFP.Cre/+};Atoh1^{A1GFP/+};R26R^{tdTomato/+}$ mice at P0 (generated by timed pregnancies and found to correspond in most cases to E19.5 and in the rest to E20.5). To determine the transcriptomes of OHCs with and without INSM1, we collected OHCs by FACS into three separate pools per genotype ($Insm1^{GFP.Cre/-}$ and $Insm1^{GFP.Cre/+}$), each with 2,200–5,000 OHCs from 8–12 E18.5 embryos. We extracted 3–7.5 ng of RNA per E18.5 OHC pool, ~3 ng per P0 IHC pool and 10–18 ng per P0 OHC pool. We used only samples with an RNA integrity number (RIN) >8.

Beijing Genomics Institute (BGI) performed sample preparation and sequencing at their facility in the Children's Hospital of Philadelphia (CHOP). The total RNA samples were first treated with DNase I to degrade any possible DNA contamination followed by ribosomal RNA removal using RiboZero (Epicentre), converted to cDNA and amplified with NuGEN Ovation RNA-Amplification System V2. Mixed with the fragmentation buffer, the mRNA was fragmented into short fragments of about 200 bp. Then the first strand of cDNA was synthesized using random hexamer-primer. Buffer, dNTPs, RNase H, and DNA polymerase I were added to synthesize the second strand. Double-stranded cDNA was purified with magnetic beads, end reparation and 3′-end single nucleotide A (adenine) addition were performed, and sequencing adaptors were ligated to the fragments, which were enriched by PCR amplification. Libraries were qualified and quantified with an Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR. Individually barcoded 100-bp paired-end library products were sequenced on the Illumina HiSeq2000 (three libraries from E18.5 $Insm1^{GFP.Cre/-}$ and three from $Insm1^{GFP.Cre/+}$ OHCs) or the HiSeq4000 (six libraries each for IHCs and OHCs from P0 $Insm1^{GFP.Cre/+};Atoh1^{A1GFP/+};R26R^{tdTomato/+}$ mice) and multiplexed per lane, yielding 48–50 million (for each of the six E18.5 OHC libraries) and 92–116 million (for each of the twelve P0 IHC or OHC libraries) paired reads. DNA read quality was evaluated in fastq format using FastQC, adapters were trimmed, and reads of poor quality or aligning to rRNA sequences were filtered.

The cleaned reads were aligned to the *Mus musculus* genome (mm10) using STAR[40]. Read counts for each gene were calculated using htseq-count[41] in conjunction with a gene annotation file for mm10 obtained from UCSC (University of California Santa Cruz; http://genome.ucsc.edu). Differential expression was determined using DESeq2[42]. The cutoff for determining significantly differentially expressed genes was an FDR-adjusted P value less than 0.05.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data are available from the corresponding authors upon reasonable request. RNA-seq data are available for public view at the gEAR portal (https://umgear.org/).

31. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
32. Kanki, H., Suzuki, H. & Itohara, S. High-efficiency CAG-FLPe deleter mice in C57BL/6J background. *Exp. Anim.* **55**, 137–141 (2006).
33. Neely, S. T. & Liu, Z. EMAV: Otoacoustic emission averager. Omaha, NE: Technical Memo No 17. *Boy's Town National Research Hospital* (1994).
34. Neely, S. T. & Stevenson, R. SysRes. Omaha, NE: Tech Memo No. 1. *BoysTown National Research Hospital NE* (1992).
35. Pearce, M., Richter, C. P. & Cheatham, M. A. A reconsideration of sound calibration in the mouse. *J. Neurosci. Methods* **106**, 57–67 (2001).
36. Cheatham, M. A. et al. Loss of the tectorial membrane protein CEACAM16 enhances spontaneous, stimulus-frequency, and transiently evoked otoacoustic emissions. *J. Neurosci.* **34**, 10325–10338 (2014).
37. Montgomery, S. C. & Cox, B. C. Whole mount dissection and immunofluorescence of the adult mouse cochlea. *J. Vis. Exp.* **107**, e53561 (2016).
38. Delmaghani, S. et al. Hypervulnerability to sound exposure through impaired adaptive proliferation of peroxisomes. *Cell* **163**, 894–906 (2015).
39. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. Staining frozen mouse embryo sections for β-galactosidase (lacZ) activity. *CSH Protoc.* **2007**, pdb.prot4726 (2007).
40. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
41. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
42. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
43. Yang, H. et al. *Gfi1-Cre* knock-in mouse line: A tool for inner ear hair cell-specific gene deletion. *Genesis* **48**, 400–406 (2010).
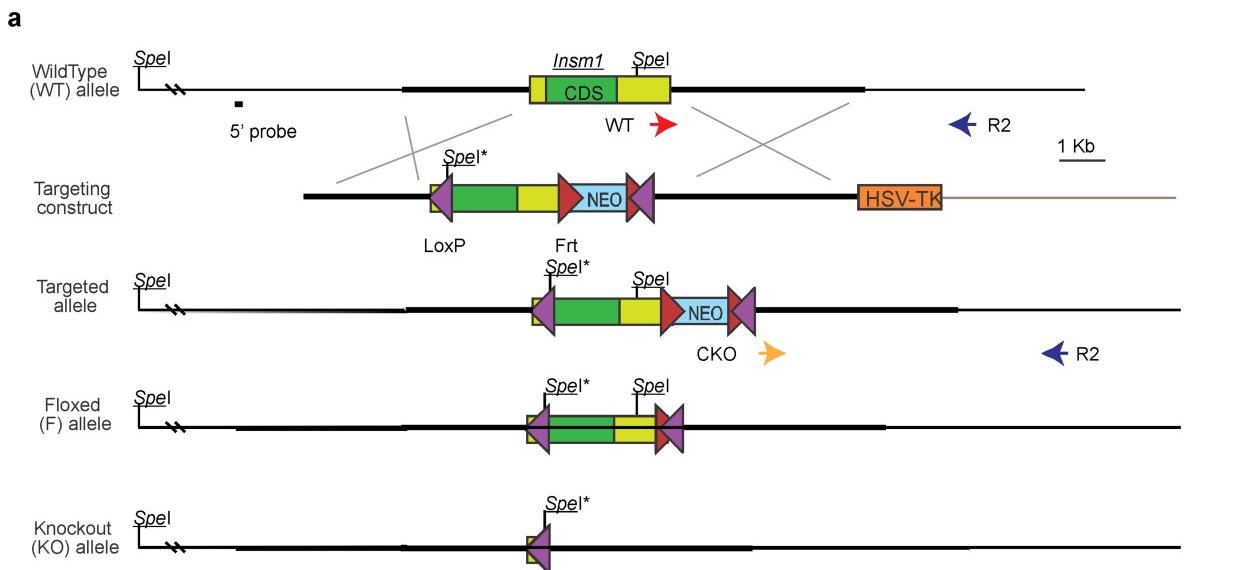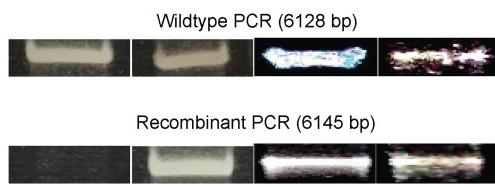
**Extended Data Fig. 1 | Generation of a conditional KO allele of *Insm1*.**
a, We generated a targeting construct in which the sole exon of *Insm1* (green rectangle, with the coding sequence in dark green and the UTRs in light green) has a *loxP* site (purple triangle) inserted in a poorly conserved area of its 5′UTR and another *loxP* site downstream of the *Insm1* gene. The construct also incorporates a neomycin resistance cassette (NEO, blue) surrounded by *Frt* sites (red triangles) and a thymidine kinase cassette (HSV-TK; orange), which are used to select for recombination events after gene targeting. **b**, We screened 439 clones and identified 5 recombinants (1 non-recombinant wild type, B5, and 3 recombinants, B6, E10 and B3, are shown) with PCR using primers indicated in **a** (arrows). The expected size for the wild-type allele using primers WT to R2 is 6,163 bp. The expected size for recombinant clones using CKO-reverse is 6,145 bp.

c, Selected embryonic stem (ES) cell clones were additionally screened for homologous recombination upstream of the first *loxP* site by Southern blotting after digestion with SpeI and using the 5′ probe indicated in **a**[6]. Southern blotting was performed twice. The expected band sizes for wild-type and conditional KO alleles are 18,162 bp and 14,333 bp, respectively. From one of these clones (B3) we generated first chimeric mice and then mice with floxed alleles of *Insm1* (obtained by crossing the chimaeras with mice expressing the FlpE recombinase (B6-Tg(CAG-FLPe)36, which deleted the NEO cassette flanked by FRT sites). Homozygous *Insm1*[F/F] mice are viable, demonstrating that the *loxP* insertions do not interfere with the vital functions of *Insm1* and hence may be used for its conditional ablation. Co-expression with Cre recombinase generates an *Insm1* KO allele lacking its entire coding sequence.
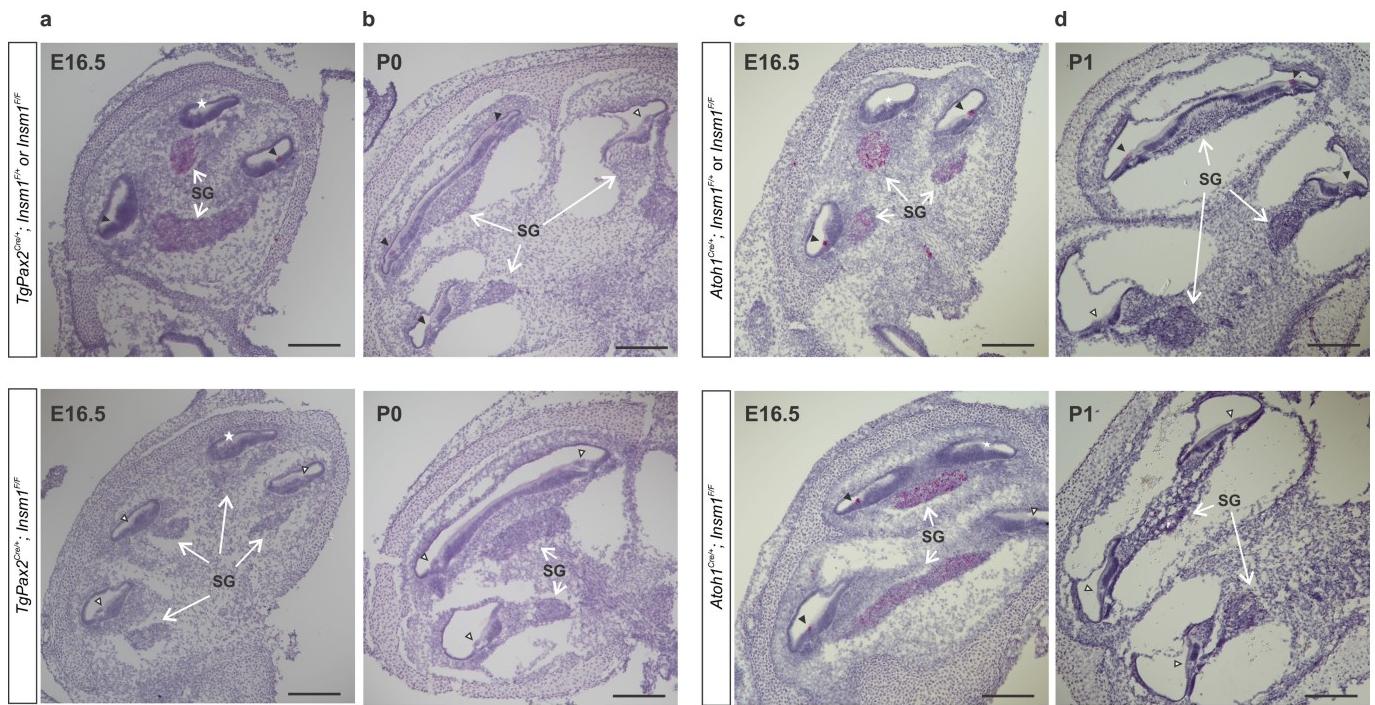
Extended Data Fig. 2 | Conditional ablation of *Insm1* in cochleae. In situ hybridization for *Insm1* transcripts on cryosections of embryonic E16.5 and neonatal (P0 and P1) cochleae. **a–c**, In control cochleae (top), *Insm1* is expressed in all OHCs (72/72 OHCs from 3 animals) and spiral ganglion (SG; white arrows) at E16.5. However, no Insm1 was detected in the organs of Corti from apical turns, in which recognizable hair cells have not yet appeared (**a**, **c**; asterisks). By postnatal age P0–P1, *Insm1* mRNA is present in 90% of OHCs (94/105 OHCs from 2 animals), and it is undetectable in spiral ganglion (**b**, **d**). **a**, **b**, Bottom, in *TgPax2^Cre/+*;*Insm1^F/F* mice, *Insm1* mRNA is undetectable in spiral ganglion and in all OHCs from E16.5 (0/69 OHCs from 2 mice; **a**) and P0 (0/42 OHCs from 1 mouse; **b**) cochleae. **c**, **d**, Bottom, in *Atoh1^Cre/+*;*Insm1^Flox/Flox* cochleae, *Insm1* mRNA is present in spiral ganglion and 43% of OHCs (18/42 OHCs from 1 animal) at E16.5 (**c**), reduced to 7% (4/54) at E17.5 (not shown), and entirely absent all OHCs (0/60 OHCs from 1 animal) and spiral ganglion by postnatal day P1 (**d**). For quantification at E16.5, we did not include organs of Corti from apical turns, which do not yet have recognizable hair cells. Filled arrowheads indicate organs of Corti with *Insm1* expression, and empty arrowheads indicate organs of Corti without *Insm1* expression. Scale bars, 200 μm.

f Calmoduluin + Myosin VIIa + DAPI
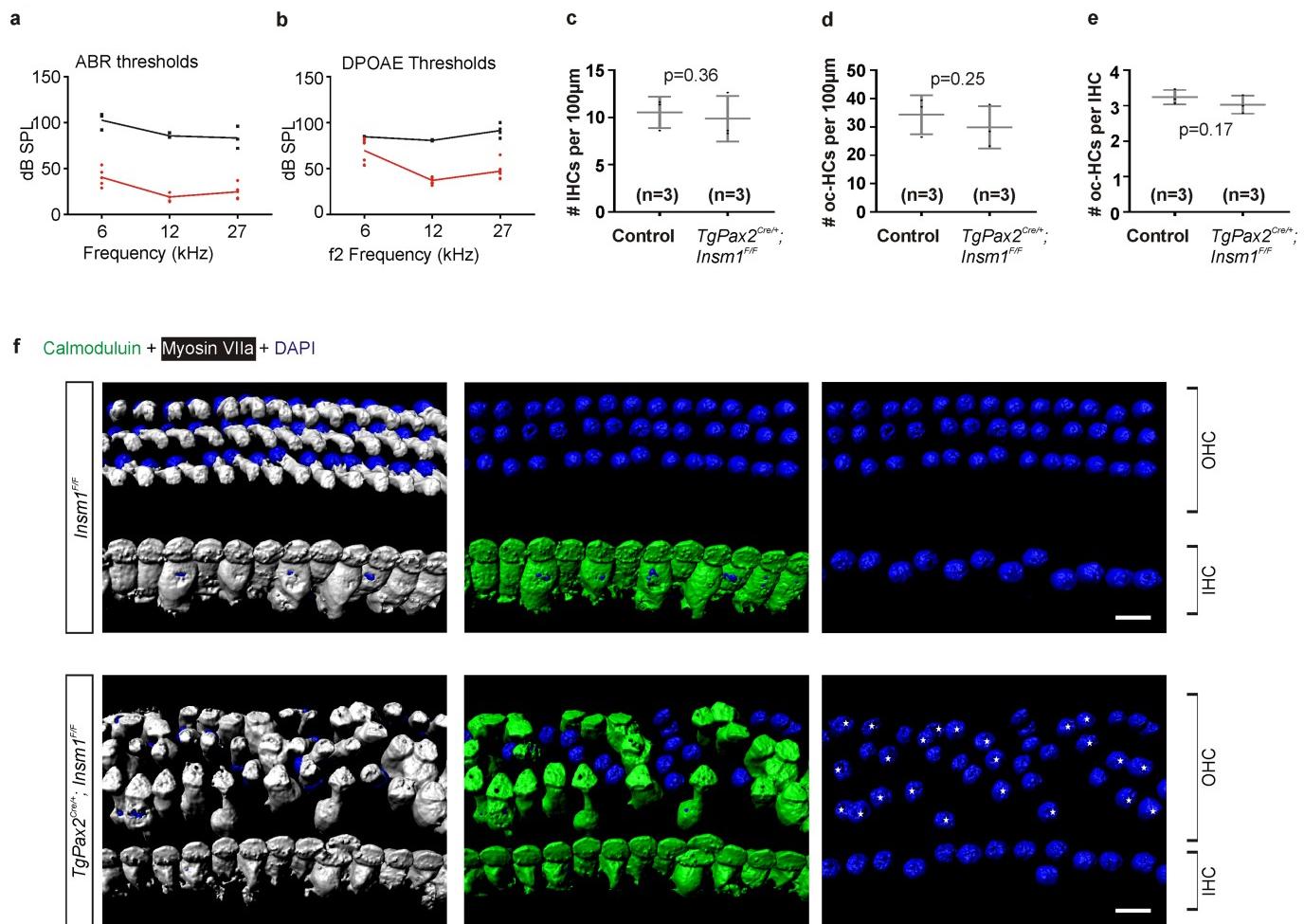
Extended Data Fig. 3 | Conditional ablation of *Insm1* in hair cells and spiral ganglia neurons using *TgPax2^Cre* causes hearing impairment and the appearance of IHC-like cells in place of OHCs.
a, b, Hearing thresholds determined by ABRs (**a**) and DPOAEs (**b**) of *TgPax2^{Cre/+}*;*Insm1^{F/F}* mice at age P35–P46 (black traces; $n = 4$, 4 females) and control littermates (red traces; $n = 5$, 2 males and 3 females). The fact that shifts in ABR threshold are larger than shifts in DPOAE threshold may indicate an additional contribution to hearing impairment of the spiral ganglion neurons lacking INSM1 in *TgPax2^{Cre/+}*;*Insm1^{F/F}* cochlea.
c, Despite the prevalence of OHCs with IHC characteristics (oc-IHCs) in *TgPax2^{Cre/+}*;*Insm1^{F/F}* cochleae (46.0 ± 5.64% (mean ± s.d.), $n = 3$ mice; Fig. 1m, n), these mice have the same density of IHCs (9.87 ± 2.41 cells per 100 μm along the organ of Corti; $n = 3$) as littermate controls (*TgPax2^{Cre/+}*;*Insm1^{F/+}* and *Insm1^{F/F}*; 10.54 ± 1.67 cells per 100 μm; $n = 3$) suggesting that oc-IHCs are not IHCs displaced from the inner to the outer compartment. **d,** There is no OHC loss in *TgPax2^{Cre/+}*;*Insm1^{F/F}* mice at ~P14–P16. Densities of oc-HCs do not differ between *TgPax2^{Cre/+}*;*Insm1^{F/F}* mice (OHCs and oc-IHCs) and their littermate

controls (OHCs only) (29.88 ± 7.45 cells per 100 μm along the organ of Corti of *TgPax2^{Cre/+}*;*Insm1^{F/F}* mice, $n = 3$; 34.3 ± 6.92 cells per 100 μm in *TgPax2^{Cre/+}*;*Insm1^{F/+}* and *Insm1^{F/F}* littermate controls, $n = 3$). **e,** The number of oc-HCs per IHC does not differ between *TgPax2^{Cre/+}*;*Insm1^{F/F}* mice and their littermate controls (3.03 ± 0.25 OHCs plus oc-IHCs per IHC in *TgPax2^{Cre/+}*;*Insm1^{F/F}* mice, $n = 3$; 3.24 ± 0.21 OHCs per IHC in *TgPax2^{Cre/+}*;*Insm1^{F/+}* and *Insm1^{F/F}* littermate controls, $n = 3$). One-tailed Student's *t*-tests were used in **c–e**, n is number of mice. Statistical significance is defined as $P < 0.05$. **f,** Immunofluorescence for the IHC-enriched calmodulin (green) and hair cell marker myosin VIIa (white) on whole-mount organs of Corti from mid-cochlear positions at ages ~P14–P16 confirmed that many *TgPax2^{Cre/+}*;*Insm1^{F/F}* oc-HCs had IHC characteristics, in addition to having the flask shape and large nuclei of IHCs (blue, DAPI, marked with asterisks), as well as lacking prestin and expressing VGLUT3 (Fig. 1m). Scale bars, 10 μm. Biological replicates were used for all experiments and similar immunohistochemistry results obtained from three or more mice per genotype.
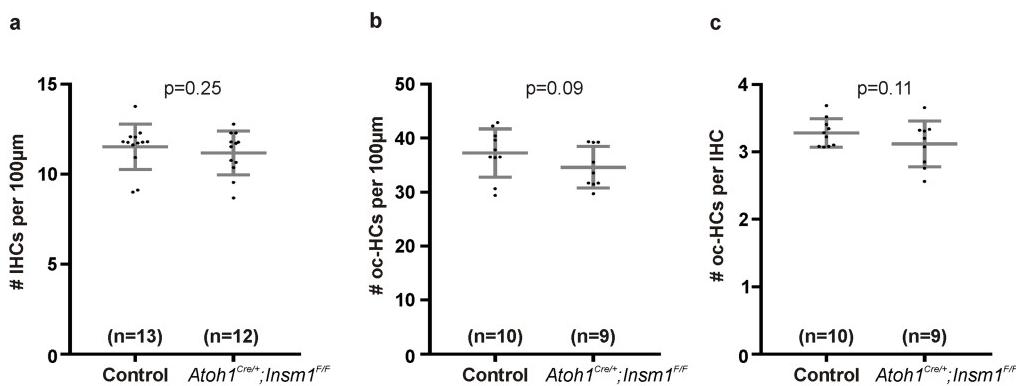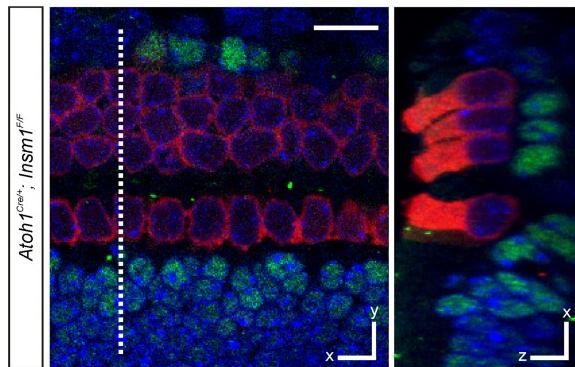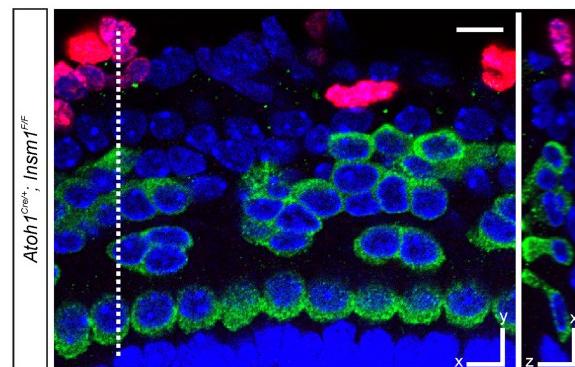
Extended Data Fig. 4 | IHC-like cells in the outer compartment (oc-IHCs) result from OHC misdifferentiation in the absence of INSM1, not from IHC displacement or from trans-differentiation of supporting cells. a, $Atoh1^{Cre/+};Insm1^{F/F}$ mice have the same density of IHCs (11.2 ± 1.2 (mean $\pm$ s.d.) cells per $100\,\mu$m; $n = 12$) as littermate controls ($Atoh1^{Cre/+};Insm1^{+/F}$ and $Insm1^{F/F}$; 11.5 ± 1.3 cells per $100\,\mu$m; $n = 13$). **b,** There is no loss of OHCs in $Atoh1^{Cre/+};Insm1^{F/F}$ mice up to P34. Densities of oc-HCs do not differ between $Atoh1^{Cre/+};Insm1^{F/F}$ mice (OHCs + oc-IHCs) and their littermate controls (OHCs only) (34.6 ± 3.8 cells per $100\,\mu$m in $Atoh1^{Cre/+};Insm1^{F/F}$ mice, $n = 9$; 37.3 ± 4.5 cells per $100\,\mu$m in $Atoh1^{Cre/+};Insm1^{+/F}$ and $Insm1^{F/F}$ littermate controls, $n = 10$). **c,** The number of oc-HCs per IHC do not differ significantly between $Atoh1^{Cre/+};Insm1^{F/F}$ mice and their littermate controls (3.1 ± 0.3 OHCs and oc-IHCs per IHC in $Atoh1^{Cre/+};Insm1^{F/F}$ mice, $n = 9$; 3.3 ± 0.2 OHCs per IHC in $Atoh1^{Cre/+};Insm1^{+/F}$ and $Insm1^{F/F}$ littermate controls, $n = 10$). The criteria for identification of oc-IHCs in $Atoh1^{Cre/+};Insm1^{F/F}$ mice were the presence in the outer compartment of hair cells expressing IHC markers (VGLUT3, high levels of calmodulin and/or nuclear CtBP2), lacking OHC markers (oncomodulin and/or prestin) and with a shape (determined

by myosin VIIa immunoreactivity) like that of IHCs. Mice used for hair cell counts were P0–P34. One-tailed Student's t-tests were used in **a–c**; n is number of mice. Statistical significance is defined as $P < 0.05$. **d,** SOX2 immunoreactivity, which labels the nuclei of cochlear supporting cells and, under certain conditions, of hair cells trans-differentiated from them postnatally, was not present in cells of the OHC region in $Atoh1^{Cre/+};Insm1^{F/F}$ pups (0/95 OHCs at P0, 0/42 OHCs at P2, and 0/39 OHCs at P5). **e,** To track postnatal cell proliferation in the organ of Corti, neonatal mice were injected twice daily with the thymidine analogue 5-ethynyl-2′-deoxyuridine (EdU) from P0 to P5 or P8. The lack of EdU in any hair cell from in $Atoh1^{Cre/+};Insm1^{F/F}$ mice (0/77 oc-HCs at P5 and 0/40 oc-HCs at P8) confirmed that these cells, including oc-IHCs, were not produced from postnatally dividing supporting cells. Unless otherwise noted, images are from mid cochlear positions. Hair cells were identified by myosin VIIa immunoreactivity, phalloidin, DAPI and Hoechst. Scale bars, $10\,\mu$m. Biological replicates were used for all experiments and similar immunohistochemistry results obtained from three or more mice per genotype.

a

ABR thresholds

b

DPOAE Thresholds

c Calmodulin + Oncomodulin
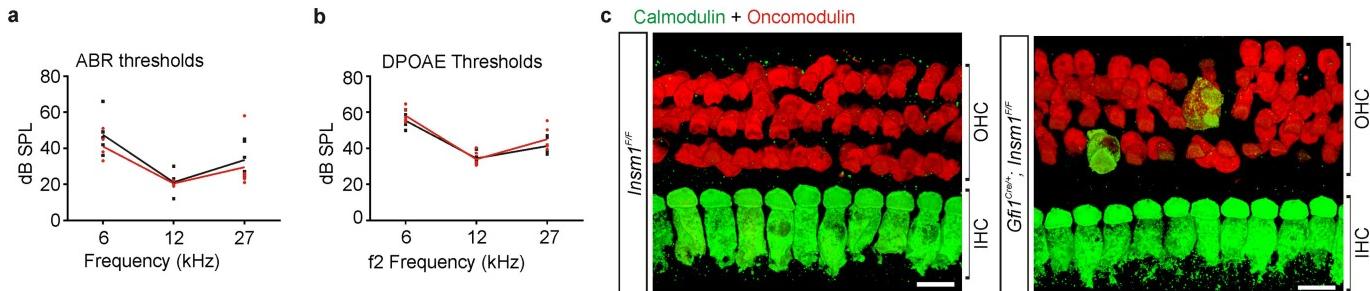
Extended Data Fig. 5 | Conditional ablation of *Insm1* in hair cells using Gfi1-Cre causes no hearing impairment, and results in only very few oc-IHCs. We generated a conditional knockout of *Insm1* in hair cells using Gfi1-Cre, in which the expression of cre recombinase coincided with that of *Insm1*. **a**, **b**, Hearing thresholds determined by ABRs (**a**) and DPOAEs (**b**) of *Gfi1^{Cre/+};Insm1^{F/F}* mice at age P30–P35 (black traces; $n = 6$, 4 males and 2 females) and control littermates (red traces; $n = 6$; 1 *Insm1^{F/F}*, 1 *Insm1^{F/+}* and 4 *Gfi1^{Cre/+};Insm1^{F/+}*; 4 males and 2 females). There is no significant difference in ABR and DPOAE thresholds at any frequency tested between *Gfi1^{Cre/+};Insm1^{F/F}* mice and their control littermates. **c**, Immunohistochemistry in whole-mount organs of Corti from mid-cochlear positions of P34 mice tested for hearing in **a**, **b** revealed that *Gfi1^{Cre/+};Insm1^{F/F}* mice had normal IHCs expressing high levels of calmodulin. However, very few oc-HCs also expressed calmodulin at high levels, oncomodulin at low levels, and had a round, flask shape similar to that of IHCs (0.78%; 5/526 OHCs from 2 mice). Because in these *Gfi1^{Cre/+};Insm1^{F/F}* mice the onset of Cre recombinase expression coincides with that of *Insm1* (E15.5–E17.5)[43], their nascent OHCs will express *Insm1* for at least several hours. This result indicates that brief expression of *Insm1* is sufficient to promote proper OHC differentiation. Scale bars, 10 μm. Biological replicates were used for all experiments and similar immunohistochemistry results obtained from three or more mice per genotype.

Extended Data Fig. 6 | FACS purification of and RNA extraction from OHCs from an E18.5 *Insm1^GFP.Cre/+* embryo. a–c, Forward and side light scattering were used to exclude dead cells and debris (**a**) and aggregates (≥2 cells) (**b**, **c**). **d**, Live cells were gated in green and red (PerCP-Cy5, to assess autofluorescence) channels to define the GFP+ (green dots) and GFP− (red dots) sorting windows. **e**, Myosin VIIa immunoreactivity and DAPI stain of cells collected through cytospinning after FACS confirm that most of the 547 sorted GFP+ cells are hair cells. This verification was done on all hair cell pools sorted (three pools per genotype). Inset is a representative merged image of one sorted OHC at high magnification. In this pool, no autofluorescent cells were collected. **f**, RT–qPCR after cell sorting (mean ± s.e.m.) reveals that, compared with GFP− cells, GFP+ cells express the hair cell marker gene *Myo7a* and not the supporting cell marker genes *S100* and *Hes5*. **g**, To ensure the quality of the extracted RNA, the RIN score was determined using a BioAnalyzer. **g**, Similar RIN scores were obtained from all pools of OHCs examined (including the three per genotype used for RNA-seq in Fig. 4a).

Extended Data Table 1 | Confirmed genes misregulated in OHCs lacking INSM1

GeneSymbol	Description	RNAseq Change (IHC/OHC)	RNAseq Change (KO/Het)	RT-qPCR Change (KO/Het)	RT-qPCR p-value (KO/Het)
Tbx2	T-box 2	19.28	13.75	102.31	0.000428
Fgf8	fibroblast growth factor 8	18.02	8.12	65.08	0.004112
Smad3	SMAD family member 3	10.69	14.04	46.43	4.14E-05
Nhs	Nance-Horan syndrome (human)	8.01	11.56	3.15	0.034775
Lrrn1	leucine rich repeat protein 1, neuronal	7.86	4.93	7.79	0.003956
Brip1	BRCA1 interacting protein C-terminal helicase 1	7.47	12.65	118.50	5.53E-05
Rin3	Ras and Rab interactor 3	5.71	10.98	1.88	0.011625
Pcdh1	protocadherin 1	4.47	3.53	2.84	0.009309
Spryd3	SPRY domain containing 3	3.64	6.70	1.53	0.033057
Pacs1	phosphofurin acidic cluster sorting protein 1	3.56	2.63	2.33	0.00229
Car13	carbonic anhydrase 13	3.55	4.27	3.69	0.003681
Tmprss7	transmembrane serine protease 7	3.03	3.07	1.58	0.033105
Rprm	reprimo, TP53 dependent G2 arrest mediator candidate	2.89	1.83	1.67	0.039198
Zfp668	zinc finger protein 668	1.85	3.23	1.30	0.044668
Mtss1	metastasis suppressor 1	1.83	2.34	1.56	0.02911
Cux1	cut-like homeobox 1	1.61	2.05	1.57	0.026945
Lrrc8b	leucine rich repeat containing 8 family, member B	1.56	1.71	1.31	0.045587
Rai1	retinoic acid induced 1	1.57	2.56	1.14	0.033262
Pink1	PTEN induced putative kinase 1	1.36	6.52	4.42	0.003253
Cmtm8	CKLF-like MARVEL transmembrane domain containing 8	1.21	8.29	1.84	0.002469
Id4	inhibitor of DNA binding 4	0.99	1.94	1.84	0.063034
Sez6l	seizure related 6 homolog like	0.23	7.47	5.04	0.020066

Differential expression of IHCs with respect to OHCs and of OHCs without INSM1 (from *Insm1*$^{GFP,Cre/-}$ mice, referred to as KO) with respect to OHCs with INSM1 (from *Insm1*$^{GFP,Cre/+}$, referred to as Het). Differential expression between KO and Het OHCs estimated by RNA-seq was confirmed by TaqMan RT–qPCR ($n = 5$ pools of OHCs per genotype for *Tbx2*, *Nhs*, *Lrrn1*, *Brip1* and *Rin3*; $n = 4$ for all other genes). *P* values are for one-tailed *t*-tests on RT–qPCR values. All 22 genes increase their expression in OHCs lacking INSM1. Of these, all except *Sez6l* are normally preferentially expressed in IHCs. Note that for *Id4*, differential expression between IHCs and OHCs was not detected by RNA-seq at P0, and it did not reach significance between KO and HET OHCs by RT–qPCR at E18.5. However, significance was achieved ($P = 0.013$) at E16.5, at which time differential expression was confirmed and visualized by RNAscope in situ hybridization (Fig. 4j). Hence, for *Id4* the differential expression occurs transiently and very early.

Helios is a key transcriptional regulator of outer hair cell maturation

Lauren Chessum[1,11], Maggie S. Matern[2,11], Michael C. Kelly[3], Stuart L. Johnson[4], Yoko Ogawa[2], Beatrice Milon[2], Mark McMurray[2], Elizabeth C. Driver[3], Andrew Parker[1], Yang Song[5], Gemma Codner[6], Christopher T. Esapa[1], Jack Prescott[1], Graham Trent[2], Sara Wells[6], Abigail K. Dragich[7], Gregory I. Frolenkov[7], Matthew W. Kelley[3], Walter Marcotti[4], Steve D. M. Brown[1], Ran Elkon[8,9], Michael R. Bowl[1,12]* & Ronna Hertzano[2,5,10,12]*

The sensory cells that are responsible for hearing include the cochlear inner hair cells (IHCs) and outer hair cells (OHCs), with the OHCs being necessary for sound sensitivity and tuning[1]. Both cell types are thought to arise from common progenitors; however, our understanding of the factors that control the fate of IHCs and OHCs remains limited. Here we identify *Ikzf2* (which encodes Helios) as an essential transcription factor in mice that is required for OHC functional maturation and hearing. Helios is expressed in postnatal mouse OHCs, and in the *cello* mouse model a point mutation in *Ikzf2* causes early-onset sensorineural hearing loss. *Ikzf2*[cello/cello] OHCs have greatly reduced prestin-dependent electromotile activity, a hallmark of OHC functional maturation, and show reduced levels of crucial OHC-expressed genes such as *Slc26a5* (which encodes prestin) and *Ocm*. Moreover, we show that ectopic expression of *Ikzf2* in IHCs: induces the expression of OHC-specific genes; reduces the expression of canonical IHC genes; and confers electromotility to IHCs, demonstrating that *Ikzf2* can partially shift the IHC transcriptome towards an OHC-like identity.

The mature mammalian cochlea contains two distinct types of sensory cells, IHCs and OHCs, each of which are highly specialized and, in humans, do not regenerate once they are damaged or lost[2]. Progressive loss of these cells, particularly the OHCs, underlies much of the aetiology of age-related hearing loss—a worldwide epidemic[3,4]. Although these two cell types were first described by Retzius in the 1800s, the mechanisms that underlie the specification of their common progenitor cells to functional inner versus outer hair cells remain poorly understood. In addition, attempts to direct stem cells towards hair cell fates have, so far, resulted only in the formation of immature cells that lack many of the markers of mature IHCs or OHCs[5]. Given the vulnerability of the OHCs, identifying factors that specify OHC fate is crucial, not only for understanding the biology of this unique cell type, but also for ultimately working towards regenerative therapies for hearing loss.

To define a set of high-confidence OHC-expressed genes for downstream gene regulation analyses, we crossed the knock-in prestin-CreER[T2] mouse, which can be induced to express Cre recombinase specifically in OHCs, with a transgenic RiboTag mouse, to enable OHC-specific ribosome immunoprecipitation[6,7]. Cochlear ducts from the resulting *RiboTag*[HA/+];*prestin*[creERT2/+] mice were collected at five postnatal time points (postnatal day (P) 8, 14 and 28, and 6 and 10 weeks), and actively translated OHC transcripts were enriched for by ribosome immunoprecipitation, followed by RNA sequencing (RNA-seq) of all immunoprecipitated and paired input RNA (Extended Data Fig. 1a, b, Supplementary Table 1). We calculated an OHC enrichment factor based on the immunoprecipitated/input RNA log2 fold change for each gene at each time point (Supplementary Table 2). Reassuringly, known

postnatal hair cell-enriched and OHC-expressed genes such as *Pou4f3*, *Gfi1*, *Strc*, *Ocm* and *Slc26a5* generally had high enrichment factor values across all time points (enrichment factor (EF) > 1), whereas prominent IHC marker genes such as *Otof*, *Atp2a3* and *Slc17a8* were generally depleted from the immunoprecipitated samples (EF < −1). In addition, marker genes for supporting cells, neurons and otic mesenchyme were also depleted (Extended Data Fig. 1c). Further informatics analyses of our RiboTag OHC dataset demonstrated a systematic enrichment of OHC markers and a depletion of IHC markers previously identified in an adult mouse OHC and IHC transcriptomic dataset[8], and classified the OHC-enriched transcripts into three clusters (Extended Data Fig. 1d–f, Supplementary Table 3). Intersecting genes with transcripts that were enriched in OHCs in our most mature RiboTag OHC data point (10 weeks, EF > 0.5) compared with the published dataset[8] resulted in a list of 100 highly confident postnatal OHC markers that are significantly and consistently enriched in postnatal OHCs (Fig. 1a, Supplementary Table 4). We and others have previously shown that relevant transcriptional regulators can be discovered by analysing the promoters of cell-type-specific genes to identify statistically over-represented transcription factor-binding motifs[9,10]. A transcription factor-binding motif prediction analysis of the 100 OHC marker genes identified several enriched motifs in the 20-kb regions that centred around the transcription start site, the top five of which correspond to the transcription factors HNF4A, MZF1, POU3F2, Helios and RFX3[11]. Of these, only *Ikzf2* (which encodes Helios) was included in the list of 100 OHC marker genes, and was found to be markedly enriched in OHCs at all time points (Fig. 1b, c), with an approximately fourfold enrichment in OHCs compared to IHCs in the previously published dataset[8] (Supplementary Table 4). Further characterization of Helios protein expression in the inner ear confirmed that it is restricted to the OHC nuclei starting from P4, and persists in functionally mature OHCs (Fig. 1d–f, Extended Data Fig. 2a). Together, these data suggest an important role for Helios in regulating the OHC transcriptome from early postnatal to adult stages.

A recent phenotype-driven *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis screen, undertaken at the MRC Harwell Institute, identified a C-to-A transversion at nucleotide 1551 of *Ikzf2* in the *cello* mouse mutant, causing a non-synonymous histidine-to-glutamine substitution (p.H517Q) in the encoded Helios transcription factor[12] (Fig. 1g, Extended Data Fig. 2b–d). A combination of in silico mutation analyses, structural 3D modelling, immunolabelling of Helios in the *cello* mutant mice, and in vitro assays predicted and validated a deleterious effect of the *cello* mutation on the ability of Helios to dimerize, without impairing its cellular localization (Fig. 1g, Extended Data Figs. 2e and 3). We further investigated the functional role of Helios in hearing by

Fig. 1 | Helios is a candidate regulator of OHC genes. a, The 100 OHC marker genes ($n = 100$) are enriched in OHCs at all RiboTag OHC dataset time points compared to the expression of all other genes detected (background, BG) ($n = 13,044$). *P* values: P8 = 1.73×10^{-17}, P14 = 6.55×10^{-12}, P28 = 1.60×10^{-18}, 6 weeks (w) = 7.79×10^{-18}, 10 weeks = 1.43×10^{-33} (two-sided Wilcoxon's test). Centre line represents median enrichment factor (EF; $\log_2$ fold change), box demarcates first and third quartiles, whiskers demarcate first and third quartile $\pm 1.5 \times$ interquartile range (IQR) values, dots denote single outliers. **b,** Transcription factor-binding motif analysis using the 100 highly confident OHC marker genes identifies the binding signature for Helios as significantly overrepresented. Normalized enrichment score (NES) = 3.85; NES ≥ 3.0 corresponds to a false discovery rate (FDR) of 3–9%; see ref. [11]. **c,** *Ikzf2* transcript enrichment in OHCs as measured by RiboTag OHC RNA-seq. **d,** Specific expression of Helios in the nuclei of wild-type P8 OHCs (white arrows). $n = 3$ biologically independent samples. Scale bar, 50 μm. OoC, organ of Corti; RM, Reissner's membrane;

SG, spiral ganglion; SL, spiral ligament; SV, stria vascularis. **e,** Helios expression is maintained in wild-type OHCs at 1 month (white arrows). $n = 3$ biologically independent samples. Scale bar, 10 μm. **f,** Helios is detected in wild-type OHCs from P4 and is maintained in mature P16 OHCs. $n = 2$ (P3) and $n = 4$ (P4, P8 and P16) biologically independent samples. Specificity is confirmed by the loss of labelling when the anti-Helios antibody is 'pre-blocked' with its immunizing peptide. $n = 5$ biologically independent samples. Scale bars, 10 μm. HC, Hensen's cells; PC, pillar cells. **g,** Top, the genomic and domain structure of *Ikzf2*. Black, 5′ untranslated region; light grey, N-terminal DNA-binding domain; dark grey, C-terminal dimerization domain. The *Ikzf2^cello* mutation lies in ZnF6. Bottom, further alignment of the Helios ZnF6 sequence with its paralogues and the classical Cys$_2$His$_2$ ZnF motif shows that the H517Q *cello* mutation causes substitution of a highly conserved zinc-coordinating histidine residue. 3D modelling of wild-type *Ikzf2^+* ZnF6 and mutant *Ikzf2^cello* ZnF6 illustrates the requirement of residue His517 for zinc coordination, which is not possible when residue Gln517 is substituted.

assessing auditory brainstem response (ABR) thresholds in wild-type and *cello* mice across several time points. Results show that *Ikzf2^cello/cello* mice have progressive deterioration of hearing function that starts as early as P16 (>60 dB sound pressure level (SPL)), with a threshold of ≥ 85 dB SPL by 9 months (Fig. 2a, b, Extended Data Fig. 4a–c). Using scanning electron microscopy, we show that the ultrastructure of the cochlear sensory epithelia and hair cell stereocilia bundles in the *cello* mice appear normal up to 1 month of age, after which the OHC bundles, and later the IHC bundles, begin to degenerate (Extended Data Figs. 4d, 5a–d, Supplementary Tables 5, 6). These data indicate that the hearing impairment in *cello* mice precedes the loss of hair cell bundles, and suggest that the Helios mutation instead leads to a functional deficit in OHCs. Furthermore, by using a second *Ikzf2* mutant allele (*Ikzf2^del890*), which leads to an in-frame deletion of the third coding exon, we confirm *Ikzf2* as the causative gene that underlies the auditory dysfunction in the *cello* mutants. At 1 month of age, *Ikzf2^cello/del890* compound heterozygotes display increased ABR thresholds (up to 40 dB SPL) compared to heterozygotes and wild-type mice (Extended Data Fig. 5e, f), confirming *Ikzf2^cello* as the causative allele in the *cello* mutant.

To explore the effect of the *cello* mutation on OHC physiology, we investigated the basolateral properties of OHCs. We found that the mechanoelectrical transducer (MET) current (Extended Data Fig. 6a–c) and the adult-like potassium (K$^+$) current $I_{K,n}$ (Extended Data Fig. 6d–h) have normal biophysical characteristics in *Ikzf2^cello/cello* OHCs. The resting membrane potential (V_m) of OHCs is also similar between genotypes (*Ikzf2^cello/+*: -68 ± 2 mV (mean $\pm$ s.e.m.); *Ikzf2^cello/cello*: -70 ± 1 mV). We then investigated whether Helios regulates OHC electromotile activity. We found that stepping the membrane potential from -64 mV to $+56$ mV causes the OHCs from both genotypes to shorten (Fig. 2c, d), as previously described[13–15]. However, *Ikzf2^cello/cello* OHCs show significantly reduced movement compared to *Ikzf2^cello/+* control OHCs (Fig. 2e), even when the values are normalized to their reduced surface area (Fig. 2f). We also found that young adult *Ikzf2^cello/cello* mice have significantly reduced distortion product oto-acoustic emission (DPOAE) responses (≤ -15 dB SPL) compared to littermate controls (Fig. 2g), further demonstrating impaired OHC function.

To identify genes regulated by Helios in OHCs, we compared gene expression from the cochleae of P8 *Ikzf2^cello/cello* and their wild-type
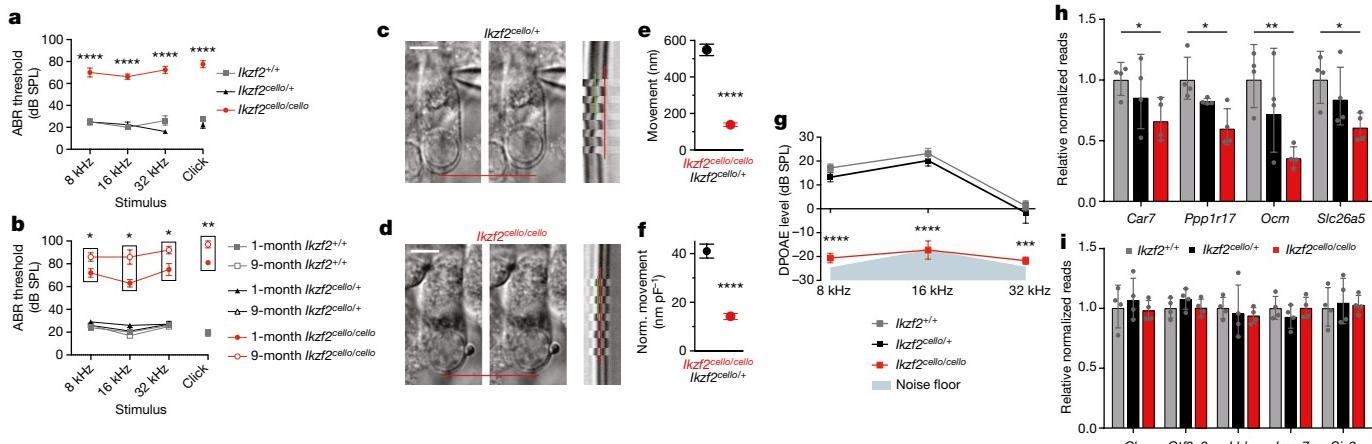
Fig. 2 | Helios is required for hearing and OHC electromotility.
a, b, Averaged ABR thresholds for *cello* mice at P16 (**a**) and 1 and
9 months of age (**b**). Age-matched *Ikzf2*$^{+/+}$ and *Ikzf2*$^{cello/+}$ controls display
thresholds within the expected range (15–30 dB SPL) at all time points
tested. $n = 4$ (**a**) and $n = 5$ (**b**) biologically independent animals per
genotype at each time period. Data are mean ± s.e.m. ****$P < 0.0001$
(P16 *Ikzf2*$^{cello/cello}$ vs *Ikzf2*$^{+/+}$ and vs *Ikzf2*$^{cello/+}$ at 8 kHz, 16 kHz, 32 kHz,
and click stimulus); *$P = 0.0284$ (1- vs 9-month *Ikzf2*$^{cello/cello}$ at 8 kHz);
*$P = 0.0166$ (1- vs 9-month *Ikzf2*$^{cello/cello}$ at 16 kHz); *$P = 0.0303$ (1- vs
9-month *Ikzf2*$^{cello/cello}$ at 32 kHz); **$P = 0.0042$ (1- vs 9-month *Ikzf2*$^{cello/cello}$
click stimulus) (one-way ANOVA with Tukey post hoc test (**a**) or
two-sided Welch's t-test (**b**)). See also Extended Data Fig. 4. **c, d**, Left,
images show a patch pipette attached to an OHC from control *Ikzf2*$^{cello/+}$
(**c**) and mutant *Ikzf2*$^{cello/cello}$ cochleae at P16–P18 (**d**). Red lines indicate
the position of the OHC basal membrane before (left) and during (right)
a depolarizing voltage step from −64 mV to +56 mV, highlighting
the shorting of the cells. Scale bars, 5 μm. Right, time-based z-stack
projections, in which red lines indicate the resting position of the basal
membrane and the green lines indicate the movement. $n = 10$ (*Ikzf2*$^{cello/+}$)

and $n = 21$ (*Ikzf2*$^{cello/cello}$) z-stack projections (one set per OHC) from 5
biologically independent animals per genotype. **e, f**, Average movement
was significantly reduced in *Ikzf2*$^{cello/cello}$ OHCs compared to *Ikzf2*$^{cello/+}$
at P16–P18 (**e**), even after normalization to respective membrane
capacitance (**f**) (for this set of recordings, *Ikzf2*$^{cello/+}$: 13.6 ± 0.4 pF;
Ikzf2$^{cello/cello}$: 10.0 ± 0.3 pF). Data are mean ± s.e.m. $n = 10$ (*Ikzf2*$^{cello/+}$) and
$n = 21$ (*Ikzf2*$^{cello/cello}$) OHCs from 5 biologically independent animals per
genotype. ****$P < 0.0001$ (two-sided Welch's t-test). **g**, Average DPOAE
responses for *cello* mice at 1 month of age ($n = 5$ biologically independent
animals per genotype). Data are mean ± s.e.m. ****$P < 0.0001$ (*Ikzf2*$^{cello/cello}$
vs *Ikzf2*$^{+/+}$ and vs *Ikzf2*$^{cello/+}$ at 8 kHz, 16 kHz); ***$P = 0.0004$ (*Ikzf2*$^{cello/cello}$
vs *Ikzf2*$^{+/+}$ at 32 kHz); ***$P = 0.0012$ (*Ikzf2*$^{cello/cello}$ vs *Ikzf2*$^{cello/+}$ at 32 kHz)
(one-way ANOVA with Tukey post hoc test). **h, i**, NanoString validations
of genes downregulated in *Ikzf2*$^{cello/cello}$ cochleae at P8 (**h**) and results
showing no change in expression of other OHC transcription factors (**i**).
Data are normalized to wild-type (*Ikzf2*$^{+/+}$) and shown as mean ± s.d.
($n = 4$ biologically independent samples per genotype). *$P = 0.028$
(*Car7*; *Ikzf2*$^{cello/cello}$ vs *Ikzf2*$^{+/+}$); *$P = 0.017$ (*Ppp17r1*); **$P = 0.006$ (*Ocm*);
*$P = 0.017$ (*Slc26a5*) (two-sided Welch's t-test).
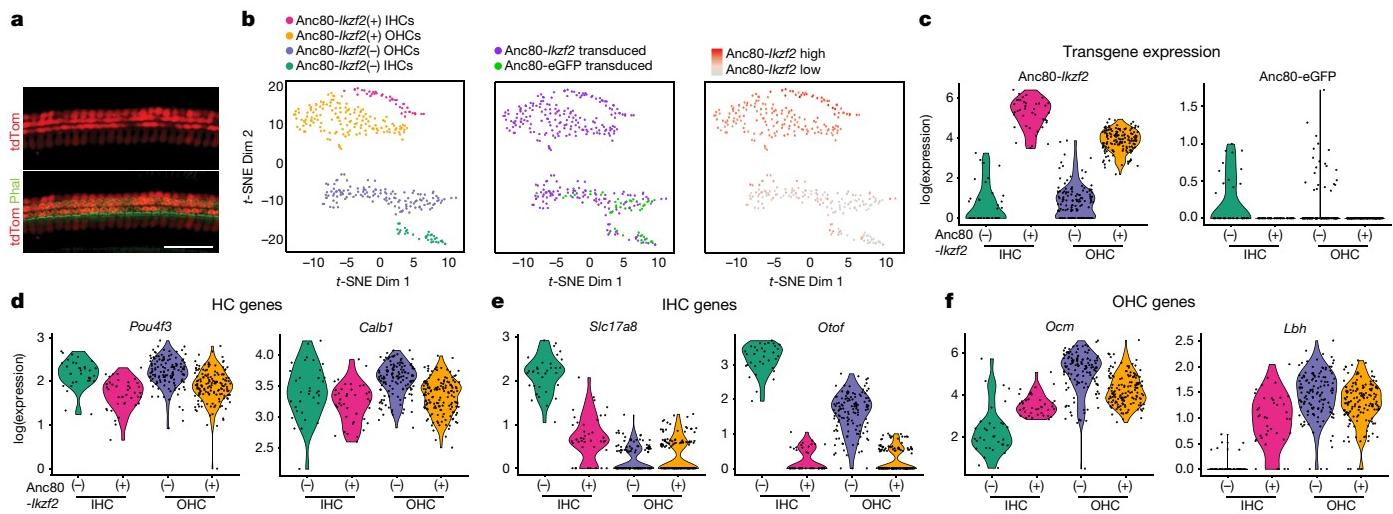
**Fig. 3 | Partial transcriptional conversion of Anc80-*Ikzf2* transduced IHCs
identified by scRNA-seq. a**, Representative *Myo15*$^{cre/+}$;*ROSA26*$^{CAG-tdTomato}$
cochlear whole-mount staining. *Myo15*^{cre}-driven tdTomato expression
is hair cell specific at P6 ($n = 3$ biologically independent samples with
similar results). Scale bar, 50 μm. **b**, t-distributed stochastic neighbour
embedding (t-SNE) plots of all cochlear hair cells profiled by scRNA-seq,
including the cluster to which each cell was assigned, the experimental
origin of each cell (cochlea injected with Anc80-*Ikzf2* or Anc80-eGFP),
and the relative transcript abundance of Anc80-*Ikzf2* measured in each
cell. **c**, Anc80-*Ikzf2* is highly expressed in the Anc80-*Ikzf2*(+) IHCs
and OHCs, whereas Anc80-eGFP expression is only seen in the cells
assigned to the Anc80-*Ikzf2*(−) IHC and OHC clusters. Dots represent the
expression values of individual cells, with width of violins summarizing

overall relative distribution of expression. **d**, Canonical hair cell (HC)
markers are highly expressed in all clusters, and not notably changed
as a result of Anc80-*Ikzf2* expression. **e**, IHC-enriched genes that are
highly expressed in control IHCs vs control OHCs, but are significantly
reduced in Anc80-*Ikzf2*(+) IHCs. Anc80-*Ikzf2*(−) IHC ($n = 34$) vs Anc80-
Ikzf2(+) IHC ($n = 40$) FDR: *Slc17a8* = 2.25×10^{-12}, *Otof* = 6.76×10^{-14}
(Kruskal–Wallis test followed by post hoc pairwise Wilcoxon ranked
sum test adjusted for multiple comparisons). **f**, OHC-enriched genes that
are induced in Anc80-*Ikzf2*(+) IHCs. Anc80-*Ikzf2*(−) IHC ($n = 34$) vs
Anc80-*Ikzf2*(+) IHC ($n = 40$) FDR: *Ocm* = 3.65×10^{-8}, *Lbh* = 1.81×10^{-10}
(Kruskal–Wallis test followed by post hoc pairwise Wilcoxon ranked
sum test adjusted for multiple comparisons). See also Extended Data
Figs. 8 and 9.
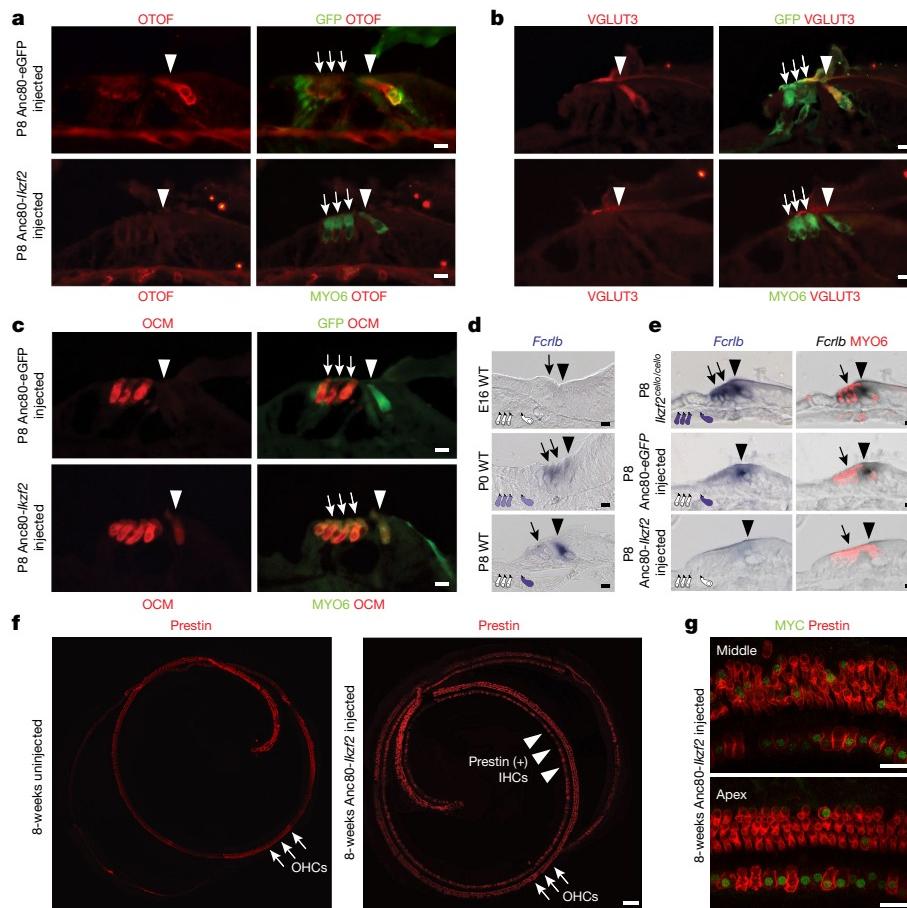
Fig. 4 | Helios overexpression modulates expression of hair cell markers. a, b, The IHC markers OTOF and VGLUT3 are downregulated in Anc80-*Ikzf2*-transduced IHCs (*n* = 3 biologically independent samples). **c,** The OHC marker OCM is expressed in Anc80-*Ikzf2*-transduced IHCs (*n* = 3 biologically independent samples per condition). **d,** *Fcrlb* expression during wild-type (WT) mouse inner ear development as detected by in situ hybridization. At embryonic day (E) 16, *Fcrlb* expression is not detected in the inner ear, but by P0 it is detected in both IHCs and OHCs, and is largely restricted to the IHCs by P8 (*n* = 3 biologically independent samples per time point). **e,** In the absence of functional Helios (*Ikzf2*^cello/cello^ mouse), *Fcrlb* is robustly expressed in IHCs and OHCs at P8. IHC expression of *Fcrlb* is not affected by Anc80-eGFP transduction, whereas *Fcrlb* expression is lost in Anc80-*Ikzf2*-transduced hair cells (*n* = 3 biologically independent samples per condition). **f, g,** Prestin expression can be seen in Anc80-*Ikzf2*-transduced IHCs up to 8 weeks of age (*n* = 3 biologically independent samples at 6–8 weeks) (**f**), and overlaps with MYC staining (**g**). Scale bars, 10 μm (**a–e**), 100 μm (**f**) and 20 μm (**g**). Arrows denote OHCs, arrowheads denote IHCs. See also Extended Data Fig. 10.

littermate controls by RNA-seq. We identified 105 upregulated and 36 downregulated genes in *Ikzf2*^cello/cello^ cochleae (Supplementary Table 7), including downregulation of the canonical OHC markers *Slc26a5* and *Ocm*, which was confirmed by NanoString validation (Fig. 2h). Furthermore, we did not observe modulation of other OHC-expressed transcription factors selected from a previously published dataset[16] (Fig. 2i), suggesting that the observed dysregulation in OHC genes results from disruption of a specific transcriptional cascade. Notably, by P16, the transcript levels of *Car7*, *Ocm* and *Slc26a5*, but not *Ppp1r17*, in *Ikzf2*^cello/cello^ cochleae are similar to the levels of wild-type littermate controls, suggesting that other factors may be compensating for the functional loss of Helios by this time point (Extended Data Fig. 6i).

To characterize the transcriptional cascade downstream of Helios, we performed in vivo Anc80L65 adeno-associated virus (AAV) gene delivery of a Myc-tagged *Ikzf2* or enhanced green fluorescent protein (eGFP) (hereafter termed Anc80-*Ikzf2* or Anc80-eGFP, respectively) to neonatal inner ears of *Myo15*^cre/+^;*ROSA26*^CAG-tdTomato^ mice, sorted the cochlear hair cells at P8, and measured resultant changes in gene expression using single-cell RNA sequencing (scRNA-seq)[17,18] (Fig. 3a, Extended Data Fig. 7). The hair cells from inner ears injected with Anc80-*Ikzf2* separated into two distinct sets of clusters, containing both IHCs and OHCs. One set of IHCs and OHCs completely overlapped with the hair cells from the control ears injected with Anc80-eGFP (Fig. 3b, bottom clusters), whereas the other set clustered separately

(Fig. 3b, top clusters). Separation of the two sets of clusters showed a clear correlation with expression of the Anc80-*Ikzf2* transgene (Fig. 3b), in which hair cells in the bottom clusters had lower expression of Anc80-*Ikzf2*, and the hair cells in the top clusters had higher expression of Anc80-*Ikzf2* (hereafter defined as Anc80-*Ikzf2* low (−) and high (+), respectively). Because the hair cells defined as Anc80-*Ikzf2*(−) clustered together with the hair cells transduced with Anc80-eGFP, these two groups of hair cells were merged and named Anc80-*Ikzf2*(−) IHCs and OHCs for all downstream analyses (Fig. 3b, c).

Although the overexpression of *Ikzf2* in IHCs and OHCs did not change the expression of hair cell markers such as *Pou4f3* and *Calb1* (Fig. 3d), it led to a significant downregulation of many genes whose transcripts were identified as IHC-enriched in the control hair cell populations, including *Slc17a8*, *Otof*, *Rprm*, *Atp2a3* and *Fgf8* (Fig. 3e, Extended Data Fig. 8, Supplementary Tables 8–10). Notably, some of the genes that are downregulated in both Anc80-*Ikzf2*(+) IHCs and OHCs are genes that are normally expressed in both cell types in early postnatal development, and that later become IHC-specific[19,20] (for example, *Pvalb* and *Otof*; Supplementary Table 10). This suggests that the overexpression of *Ikzf2* in OHCs results in an accelerated downregulation of these genes. Furthermore, *Ikzf2* overexpression in IHCs results in the upregulation of genes that are normally enriched in OHCs, such as *Ocm*, *Pde6d*, *Ldhb* and *Lbh* (Fig. 3f, Extended Data Fig. 8). Overall, these data suggest that during normal OHC development, Helios

functions to decrease the expression of early pan-hair-cell markers, such as *Otof*, in the maturing OHCs, as well as to upregulate OHC marker genes. A correlation analysis further validates the role of Helios in regulating OHC-related gene expression (Extended Data Figs. 8, 9, Supplementary Table 11). The effect of *Ikzf2* transduction on IHC gene expression was also validated by immunolabelling for OTOF, VGLUT3, OCM or prestin and by in situ hybridization for *Fcrlb* (Fig. 4, Extended Data Fig. 10a, b). Analysis of the surface characteristics of the transduced IHCs does not show a change from an IHC-like to an OHC-like stereociliary bundle, consistent with a partial role for Helios in regulating OHC-fate (Extended Data Fig. 10c). However, *Ikzf2* transduction resulted in the appearance of prominent voltage-dependent (nonlinear) capacitance in IHCs (Extended Data Fig. 10d, e), which is an electrical 'signature' of prestin-dependent OHC electromotility[21,22]. These data indicate that Anc80-*Ikzf2*-transduced IHCs start to acquire the major function of normal OHCs.

In conclusion, our study demonstrates that Helios is necessary for hearing and is a crucial regulator of gene expression in the maturing postnatal OHC. In particular, our results suggest that Helios functions to suppress IHC and early pan-hair-cell gene expression in OHCs, as well as to upregulate canonical OHC marker genes. It further shows that Helios is sufficient to induce the essential functional characteristic of electromotility and many of the molecular characteristics of OHCs when expressed in early postnatal IHCs, albeit not all of them, supporting the notion that additional OHC-expressed transcription factors are involved in postnatal OHC development. To our knowledge, this is the first study to demonstrate functional shifts in postnatal hair-cell molecular identities via viral gene delivery, and it suggests that the delivery of combinations of transcription factors may lead to successful regeneration of functional OHCs in the deafened cochlea.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0728-4.

1. Dallos, P. et al. Prestin-based outer hair cell motility is necessary for mammalian cochlear amplification. *Neuron* **58**, 333–339 (2016).
2. Ehret, G. Development of absolute auditory thresholds in the house mouse (*Mus musculus*). *J. Am. Audiol. Soc.* **1**, 179–184 (1976).
3. Bielefeld, E. C., Tanaka, C., Chen, G. D. & Henderson, D. Age-related hearing loss: is it a preventable condition? *Hear. Res.* **264**, 98–107 (2010).
4. World Health Organization. *Deafness and Hearing Loss*; available at: http://www.who.int/mediacentre/factsheets/fs300/en/ (2018).
5. Mittal, R. et al. Recent advancements in the regeneration of auditory hair cells and hearing restoration. *Front. Mol. Neurosci.* **10**, 236 (2017).
6. Fang, J. et al. Outer hair cell-specific prestin-CreER[T2] knockin mouse lines. *Genesis* **50**, 124–131 (2012).
7. Sanz, E. et al. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc. Natl Acad. Sci. USA* **106**, 13939–13944 (2009).
8. Liu, H. et al. Characterization of transcriptomes of cochlear inner and outer hair cells. *J. Neurosci.* **34**, 11085–11095 (2014).
9. Elkon, R. et al. RFX transcription factors are essential for hearing in mice. *Nat. Commun.* **6**, 8549 (2015).
10. Hertzano, R. et al. Cell type-specific transcriptome analysis reveals a major role for Zeb1 and miR-200b in mouse inner ear morphogenesis. *PLoS Genet.* **7**, e1002309 (2011).
11. Janky, R. et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLOS Comput. Biol.* **10**, e1003731 (2014).
12. Potter, P. K. et al. Novel gene function revealed by mouse mutagenesis screens for models of age-related disease. *Nat. Commun.* **7**, 12444 (2016).
13. Ashmore, J. F. A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier. *J. Physiol. (Lond.)* **388**, 323–347 (1987).
14. Brownell, W. E., Bader, C. R., Bertrand, D. & de Ribaupierre, Y. Evoked mechanical responses of isolated cochlear outer hair cells. *Science* **227**, 194–196 (1985).
15. Marcotti, W. & Kros, C. J. Developmental expression of the potassium current $I_{K,n}$ contributes to maturation of mouse outer hair cells. *J. Physiol. (Lond.)* **520**, 653–660 (1999).
16. Li, Y. et al. Transcription factors expressed in mouse cochlear inner and outer hair cells. *PLoS One* **11**, e0151291 (2016).
17. Zinn, E. et al. In silico reconstruction of the viral evolutionary lineage yields a potent gene therapy vector. *Cell Reports* **12**, 1056–1068 (2015).
18. Landegger, L. D. et al. A synthetic AAV vector enables safe and efficient gene transfer to the mammalian inner ear. *Nat. Biotechnol.* **35**, 280–284 (2017).
19. Roux, I. et al. Otoferlin, defective in a human deafness form, is essential for exocytosis at the auditory ribbon synapse. *Cell* **127**, 277–289 (2006).
20. Simmons, D. D., Tong, B., Schrader, A. D. & Hornak, A. J. Oncomodulin identifies different hair cell types in the mammalian inner ear. *J. Comp. Neurol.* **518**, 3785–3802 (2010).
21. Santos-Sacchi, J. Reversible inhibition of voltage-dependent outer hair cell motility and capacitance. *J. Neurosci.* **11**, 3096–3110 (1991).
22. Zheng, J. et al. Prestin is the motor protein of cochlear outer hair cells. *Nature* **405**, 149–155 (2000).

Author contributions L.C., M.S.M., M.C.K., Y.O., B.M., M.M., G.C., C.T.E., G.I.F., M.W.K., W.M., S.D.M.B., M.R.B. and R.H. designed and interpreted the experiments. L.C., M.S.M., M.C.K., S.L.J., Y.O., B.M., M.M., A.P., J.P., R.H., E.C.D., G.T., A.K.D., G.I.F. and W.M. performed the experiments. Y.S. and R.E. analysed the gene expression data. S.W. aided in the management of the *cello* colony. M.R.B. and R.H. conceived and coordinated the study. L.C., M.S.M., M.C.K., W.M., S.D.M.B., R.E., M.R.B. and R.H. wrote the manuscript.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment unless stated otherwise.

Animal procedures. Animal procedures performed at the University of Maryland School of Medicine were carried out in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and have been approved by the Institutional Animal Care and Use Committee at the University of Maryland, Baltimore (protocol numbers 1112005 and 1015003). The RiboTag (maintained on a C57BL/6N background), $prestin^{creERT2}$ and $Myo15^{cre}$ mouse models (maintained on a C57BL/6J background) have been described previously[6,7,23], and were provided by M. K. Lobo (RiboTag), J. Zuo ($prestin^{creERT2}$), and C. Petit and T. Friedman ($Myo15^{cre}$). CBA/CaJ mice (stock 000654) and B6.Cg-$Gt(ROSA)26Sor^{tm14(CAG-tdTomato)Hze}$/J mice (stock 007914, referred to as $ROSA26^{CAG-tdTomato}$) were procured from the Jackson Laboratory. The specificity of $prestin^{creERT2}$ was determined by crossing $prestin^{creERT2/creERT2}$ mice to $ROSA26^{CAG-tdTomato}$ mice, and the resulting offspring was dissected at P21 for whole-mount immunohistochemistry. To generate animals for the RiboTag OHC RNA-seq dataset, $RiboTag^{HA/HA}$ mice were crossed to $prestin^{creERT2/creERT2}$ mice to produce $RiboTag^{HA/+};prestin^{creERT2/+}$. These mice were further intercrossed to obtain double homozygous $RiboTag^{HA/HA};prestin^{creERT2/creERT2}$ animals, which were then crossed to CBA/CaJ mice to generate F_1 $RiboTag^{HA/+};prestin^{creERT2/+}$ offspring on a mixed CBA/C57BL/6 background, avoiding the recessively inherited age-related hearing loss phenotype inherent to C57BL/6 mice[24]. Recombination was induced by tamoxifen injection (3 mg per 40 g body weight in mice younger than 21 days, 9 mg per 40 g body weight in mice 21 days or older), and cochlear tissues were collected at the following ages: P8, P14, P28, 6 weeks and 10 weeks. For the cello RNA-seq and NanoString experiments, cochlear ducts from $Ikzf2^{+/+}$, $Ikzf2^{cello/+}$ and $Ikzf2^{cello/cello}$ mice were dissected at P8 and P16. CD-1 or C57BL/6 pregnant females were procured from Charles River or the University of Maryland School of Medicine Veterinary Resources. Resulting neonates were injected with Anc80L65 virus between P1 and P3, and dissected for later analyses between P8 and 8 weeks. For the Anc80L65-transduced IHC scRNA-seq experiment, $Myo15^{cre/cre}$ mice were crossed to $ROSA26^{CAG-tdTomato}$ mice, and resulting offspring was injected with Anc80L65 virus between P1 and P3, and the cochlear epithelium was collected at P8. Additionally, several litters with Anc80-$Ikzf2$-injected pups and their control littermates (aged P7–P8), together with a mother, were sent to the University of Kentucky for the measurements of nonlinear (voltage-dependent) capacitance, an electrical 'signature' of electromotility. All animal procedures for these experiments were approved by the Institutional Animal Care and Use Committee at the University of Kentucky (protocol 00903M2005). Both male and female animals were used for all experiments.

Animal procedures performed at the MRC Harwell Institute were licensed by the Home Office under the Animals (Scientific Procedures) Act 1986, UK and additionally approved by the relevant Institutional Ethical Review Committees. The cello mutant mouse was originally identified from the MRC Harwell Institute phenotype-driven N-ethyl-N-nitrosourea (ENU) Ageing Screen[12]. In this screen, ENU-mutagenized C57BL/6J males were mated with wild-type 'sighted C3H' (C3H.Pde6b+) females[25]. The resulting G_1 males were crossed with C3H. Pde6b+ females to produce G_2 females, all of which were screened for the $Cdh23^{ahl}$ allele[24]. $Cdh23^{+/+}$ G_2 females were then backcrossed to their G_1 fathers to generate recessive G_3 pedigrees, which entered a longitudinal phenotyping pipeline. Auditory phenotyping comprised clickbox testing at 3, 6, 9 and 12 months of age and ABR at 9 months of age. The $Ikzf2^{del890}$ mutant line was generated by the Molecular and Cellular Biology group at the MRC Harwell Institute using a CRISPR–Cas9-mediated deletion approach. Both male and female mice were used for experiments.

RiboTag immunoprecipitations. RiboTag immunoprecipitations were performed as described previously[7]. In brief, for one biological sample, 10 cochlear ducts from 5 mice were pooled and homogenized in 1 ml of supplemented homogenization buffer (50 mM Tris-HCl, pH 7, 100 mM KCl, 12 mM MgCl$_2$, 1% Nonidet P-40, 1 mM 1,4-dithiothreitol, 1× protease inhibitor cocktail, 200 U ml^{-1} RNaseOUT, 100 µg ml^{-1} cycloheximide, 1 mg ml^{-1} heparin). Homogenates were spun down (9,400g for 10 min at 4 °C) to remove particulates. Then, 40 µl of homogenate was reserved for total RNA isolation (input control), and the remaining homogenate was incubated with 5 µg haemagglutinin (HA) antibody (BioLegend) at 4 °C under gentle rotation for 4–6 h. The supernatant was then added to 300 µl of rinsed Invitrogen Dynabeads Protein G magnetic beads (Thermo Fisher), and incubated overnight at 4 °C under gentle rotation. The next day, bound beads were rinsed three times with 800 µl high-salt buffer (50 mM Tris-HCl, pH 7, 300 mM KCl, 12 mM MgCl$_2$, 1% Nonidet P-40, 1 mM 1,4-dithiothreitol, 100 µg ml^{-1} cycloheximide) at 4 °C for 10 min, rotating. Buffer RLT (350 µl) from the RNeasy Plus Micro kit (Qiagen) was then added to the beads or reserved input sample, and vortexed for 30 s to release bound ribosomes and RNA. RNA was extracted

according to the manufacturer's instructions for the RNeasy Plus Micro kit (Qiagen), using 16 µl of nuclease free water for elution as described previously[10]. This method yielded an average of 10.9 ng of immunoprecipitated RNA (average concentration = 0.68 ng µl^{-1}) and 185.6 ng of input RNA (average concentration = 10.9 ng µl^{-1}) for downstream analyses. All RNA samples used for RNA-seq had a minimum RNA integrity number (RIN) of 8.

cello cochlear RNA extractions. For the cello RNA-seq, cochlear ducts from P8 $Ikzf2^{+/+}$ and $Ikzf2^{cello/cello}$ mice were dissected and pooled (6 cochlear ducts per sample) to generate two biological replicates per genotype. For the NanoString validations, cochlear ducts from P8 $Ikzf2^{cello/cello}$, $Ikzf2^{cello/+}$ and $Ikzf2^{+/+}$ mice were dissected and pooled (2–4 cochlear ducts per sample) to generate four biological replicates per genotype. RNA was extracted using the Direct-zol RNA MiniPrep kit (Zymo Research) following the manufacturer's instructions. RNA quality and concentration were assessed using the Agilent RNA Pico kit (Agilent Technologies). All RNA samples used for RNA-seq had a minimum RNA integrity number (RIN) of 8.

RNA-seq and normalization. RiboTag OHC RNA-seq libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs), and samples were sequenced in at least biological duplicates on a HiSeq 4000 system (Illumina) using a 75-bp paired end read configuration. P8 $Ikzf2^{+/+}$ and $Ikzf2^{cello/cello}$ RNA libraries were prepared using the TruSeq RNA Sample Prep kit (Illumina), and samples were sequenced in biological duplicates on a HiSeq 2000 system (Illumina) and a 125-bp paired-end read configuration. Reads were aligned to the Mus musculus reference genome (assembly GRCm38.87 (RiboTag) or GRCm38.84 (P8 cello)) using TopHat v.2.0.8[26], and HTSeq was used to quantify the number of reads aligning to predicted coding regions[27]. See Supplementary Table 12 for alignment statistics. Expression levels were normalized using quantile normalization. In downstream analyses, only genes covered by at least 20 reads in a minimum of two samples from the same biological condition were considered as expressed. Significant differential gene expression between samples was assessed using DEseq[28]. In addition to statistical significance between samples (FDR ≤ 0.05), we also required a complete separation of expression levels between compared conditions for a gene to be called as differentially expressed. That is, for a gene to be called downregulated in condition A compared to condition B, we required that all normalized expression levels measured in the samples of condition A to be lower than all normalized expression levels measured in the samples of condition B. To avoid inflation of fold change estimates for lowly expressed genes, a floor level equal to the tenth percentile of the distribution of the expression levels was applied (that is, all expression values below the tenth percentile were set to the tenth percentile value). The OHC enrichment factors were calculated for each gene and time point by comparing the RiboTag immunoprecipitated samples to matched input samples, and are defined as the log$_2$ ratio of expression levels between the immunoprecipitated and input samples. Inspection of these enrichment factors revealed a systematic association to transcripts length (Supplementary Fig. 2a). Therefore, we used a locally weighted regression, implemented by the R lowess function, to remove this systematic effect (Supplementary Fig. 2b).

Gene expression analyses. Genes with a changed level of expression in OHC immunoprecipitated samples at any time point relative to P8 were subjected to a clustering analysis using the CLICK algorithm, implemented in the EXPANDER package[29,30]. Gene Ontology (GO) enrichment analysis was carried out using the EXPANDER implemented tool TANGO[29]. The adult mouse IHC and OHC transcriptomic dataset used for comparisons was generated previously[8] and can be accessed through the GEO database (accession number GSE111348)[8]. The expanded motif prediction analysis was performed using iRegulon[11] through the Cytoscape visualization tool[31]. The analysis was performed on the putative regulatory region of 20 kb centred around the transcription start site using default settings.

Immunohistochemistry. For cochlear sections, mice were euthanized by cervical dislocation and inner ears fixed in 4% paraformaldehyde (PFA) overnight at 4 °C then decalcified in 4% EDTA in PBS. Ears were positioned in 4% low melting temperature agarose (Sigma-Aldrich) in upturned BEEM capsules (Agar Scientific) at a 45° diagonal angle, with the apex of the cochlea facing down and the vestibular system uppermost. Once set, the agarose block was removed from the BEEM capsule and 200 µm sections were cut through the mid-modiolar plane of the cochlea using a Leica VT1000S Vibratome. Sections were simultaneously permeabilized and blocked with 10% donkey serum (Sigma) in 0.3% Triton-X for 30 min at room temperature then labelled with primary antibodies for 3 h at room temperature. To enable detection, samples were incubated with fluorophore-coupled secondary antibodies for 2 h at room temperature then stained with 4′,6-diamidino-2-phenylindole dihydrochloride (DAPI; 1:2,500, Thermo Fisher) for 5 min. Sections were transferred to WillCo glass bottom dishes (Intracel) and visualized free-floating in PBS using a Zeiss 700 inverted confocal microscope (10–40× magnification). Primary antibodies: goat anti-Helios M-20 (1:400, Santa Cruz Biotechnology) and mouse anti-β-actin (1:500, Abcam). Secondary antibodies: Alexa Fluor 568 donkey

anti-goat (Invitrogen, 1:200) and Alexa Fluor 488 donkey anti-mouse (Invitrogen, 1:200).

For cochlear whole-mounts, mice were euthanized by cervical dislocation and inner ears fixed in 2% PFA for 30 min at 4 °C. After fixation, ears were fine dissected to expose the sensory epithelium then immediately permeabilized in 0.2% Triton-X for 10 min and blocked with 10% donkey serum (Sigma) for 1 h at room temperature. Cochleae were immunolabelled with goat anti-Helios M-20 (1:400, Santa Cruz Biotechnology) overnight at 4 °C then incubated with Alexa Fluor 568 donkey anti-goat secondary (1:200, Invitrogen) and the F-actin marker Alexa Fluor 488 Phalloidin (1:200, Invitrogen) for 1 h at room temperature. Samples were washed with DAPI (1:2,500, Thermo Fisher) for 60 s to stain nuclei then mounted onto slides with SlowFade Gold (Life Technologies) and visualized using a Zeiss LSM 710 fluorescence confocal microscope and 63× oil magnification.

Identification of the *cello* mutation. DNA was extracted from ear biopsies of affected G_3 mice using the DNeasy Blood and Tissue Kit (Qiagen) and used for an initial genome-wide linkage study, using SNP markers polymorphic between the parental strains C57BL/6J and C3H.Pde6b+ (Tepnel Life Sciences). Following linkage to a 21.57 Mb region on chromosome 1, additional SNP markers were identified and genotyped using standard PCR and restriction endonuclease protocols to delineate an 8.4 Mb critical interval between SNPs rs31869113 and rs13475914. Subsequently, high-quality DNA was extracted from the tail of an affected G_3 mouse using the Illustra Nucleon BACC2 Genomic DNA Extraction Kit (GE Healthcare) and sequenced by the Oxford Genomics Centre (Wellcome Trust Centre for Human Genetics) using the HiSeq system (Illumina). Sequencing reads were aligned to the mouse reference genome (assembly GRCm38) and known C57BL/6J and C3H.Pde6b+ SNPs were filtered out, leaving variants that were then given a quality score based on their sequencing read depth. Variants within the 8.4 Mb critical region which were deemed heterozygous, low-confidence (quality score < 200), non-coding or synonymous were discounted. The putative *Ikzf2* lesion was amplified by standard PCR (see Supplementary Table 13 for genotyping primers) and validated by Sanger sequencing, using DNA from an affected G_3 animal, as well an unaffected G_3 (control). Sequence gaps that spanned coding regions were amplified by PCR using DNA from an affected G_3 mouse and analysed by Sanger sequencing. In all cases, sequence data were assessed for variation using DNASTAR Lasergene software (version 12.0.0).

In silico analyses. Three independent online tools were used to predict the functional effect of the *cello* mutation in silico: Sorting Intolerant From Tolerant (SIFT); Polymorphism Phenotyping version 2 (PolyPhen-2); and Protein Variation Effect Analyser (PROVEAN)[32–34]. Structural 3D representations of wild-type and H517Q helios ZnF6 were predicted with RaptorX[35], using peptide sequences as input, and visualized using pyMOL software (version 1.7).

In vitro analyses. A full-length *Ikzf2*+ Helios construct was prepared using the pGEM-T Vector System II Kit (Promega) and used as a template for the generation of an *Ikzf2cello* Helios construct with the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies). Plasmid DNA was prepared using the Wizard Plus SV Miniprep Purification System (Promega) and validated by Sanger sequencing. Sequence-verified *Ikzf2*+ and *Ikzf2cello* constructs were subcloned in-frame into pCMV-Myc and pEGFP-C3 mammalian expression vectors (provided by C. Esapa), to yield N-terminally tagged *Ikzf2*+ and *Ikzf2cello* Helios. See Supplementary Table 13 for cloning and mutagenesis oligonucleotide sequences.

Constructs were subsequently used for subcellular localization studies using male *Cercopithecus aethiops* SV40 transformed kidney cells (Cos-7) cells that had been seeded onto 22 × 22 mm glass coverslips in six-well plates, at a volume of 1×10^5 cells per well. After 24 h (or when 50–60% confluent), cells were transiently transfected with 1 μg DNA of the *Ikzf2*+-Myc or *Ikzf2cello*-Myc Helios construct using JetPEI DNA Transfection Reagent (Polyplus Transfection). At 24 h after transfection, cells were fixed in 4% PFA for 10 min and permeabilized with 1% Triton-X for 15 min at room temperature. After blocking in 10% donkey serum (Sigma) for 1 h at room temperature, cells were immunolabelled with goat anti-Helios M-20 primary antibody (1:600, Santa Cruz Biotechnology) overnight at 4 °C, then incubated with Alexa Fluor 488 donkey anti-goat secondary antibody (1:200, Invitrogen) and F-actin marker Texas Red-X Phalloidin (1:200, Invitrogen) for 1 h at room temperature. Cells were washed with DAPI (1:2,500, Thermo Fisher) for 60 s. Coverslips were mounted onto slides with SlowFade Gold (Life Technologies) and cells were visualized using a Zeiss LSM 710 multiphoton fluorescence confocal microscope and 63× oil magnification.

Constructs were also used for co-immunoprecipitation studies using human embryonic kidney (HEK293T) cells that had been seeded directly onto six-well plates at a volume of 5×10^5 cells per well. Cells were transiently co-transfected 24 h later with a total of 2 μg plasmid DNA to mimic the wild-type (1 μg *Ikzf2*+-Myc Helios + 1 μg *Ikzf2*+-GFP Helios), heterozygous (1 μg *Ikzf2*+-Myc Helios + 1 μg *Ikzf2cello*-GFP Helios; 1 μg *Ikzf2cello*-Myc Helios + 1 μg *Ikzf2*+-GFP Helios) or homozygous (1 μg *Ikzf2cello*-Myc Helios + 1 μg *Ikzf2cello*-GFP Helios) states using JetPEI DNA Transfection Reagent (Polyplus Transfection). Single transfections

with either 1 μg *Ikzf2*+-GFP Helios or 1 μg *Ikzf2*+-Myc Helios were also carried out for negative controls. Cells were lysed in 250 μl of 1× RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% deoxycholate, 0.1% SDS, 50 mM Tris, pH 7.5, in milliQ water) 48 h after transfection, then incubated with Protein G Sepharose Beads (Sigma) for 2 h at 4 °C. The beads were pelleted by centrifugation and the supernatant incubated with either 1 μg of mouse anti-cMyc 9E10 antibody (Developmental Studies Hybridoma Bank) or 1–2 μg of custom-made rabbit anti-GFP antibody overnight at 4 °C. The immunoprecipitation complexes were captured using Protein G beads, washed with RIPA buffer and released by incubation with NuPAGE Reducing Agent (Novex). Immunoprecipitation reactions and their corresponding reduced cell lysate were analysed by western blotting. Samples were electrophoresed on NuPage 4–12% Bis-Tris gels (Invitrogen) and transferred onto nitrocellulose membranes using the iBlot system (Invitrogen). Membranes were incubated with mouse anti-cMyc 9E10 antibody (1:5,000, Developmental Studies Hybridoma Bank) and custom-made rabbit anti-GFP (1:1,000, CUK-1819 MGU-GFP-FL) primary antibodies. Mouse 12G10 anti-α-tubulin (1:10,000, Developmental Studies Hybridoma Bank) was also used as a loading control. For detection, membranes were incubated with goat anti-mouse IRDye 680RD (1:15,000, LI-COR) and goat anti-rabbit IRDye 800CW secondary antibodies (1:15,000, LI-COR) and imaged using the Odyssey CLx Infrared Imaging System (LI-COR). For quantification, band intensities were determined using the Image Studio Lite Ver 5.2 software and used to calculate the relative ratio of the co-immunoprecipitation to immunoprecipitation signal. Cos-7 and HEK293T cell lines used in this study were provided by C. Esapa, were not authenticated, but were tested and confirmed to be free of mycoplasma contamination. Cells were grown at 37 °C under 5% CO_2 conditions in DMEM (Invitrogen) containing 10% heat-inactivated fetal bovine serum (FBS) (Invitrogen) and 1× penicillin/streptomycin (Invitrogen).

ABR. ABR tests were performed using a click stimulus in addition to frequency-specific tone-burst stimuli to screen mice for auditory phenotypes and investigate auditory function[36]. Mice were anaesthetized by intraperitoneal injection of ketamine (100 mg ml^{-1} at 10% v/v) and xylazine (20 mg ml^{-1} at 5% v/v) administered at the rate of 0.1 ml per 10 g body mass. Animals were placed on a heated mat inside a sound-attenuated chamber (ETS Lindgren) and electrodes were placed subdermally over the vertex (active), right mastoid (reference) and left mastoid (ground). ABR responses were collected, amplified and averaged using TDT System 3 (Tucker Davies Technology) in conjunction with either BioSig RP (version 4.4.11) or BioSig RZ (v5.7.1) software. The TDT system click ABR stimuli comprised clicks of 0.1 ms broadband noise spanning approximately 2–48 kHz, presented at a rate of 21.1 s^{-1} with alternating polarity. Tone-burst stimuli were of 7 ms duration, inclusive of 1 ms rise/fall gating using a Cos2 filter, presented at a rate of 42.5 s^{-1} and were measured at 8, 16 and 32 kHz. All stimuli were presented free-field to the right ear of the mouse, starting at 90 dB SPL and decreasing in 5 dB increments. Auditory thresholds were defined as the lowest dB SPL that produced a reproducible ABR trace pattern and were determined manually. All ABR waveform traces were viewed and re-scored by a second operator blinded to genotype. Animals were recovered using 0.1 ml of anaesthetic reversal agent atipamezole (Antisedan, 5 mg ml^{-1} at 1% v/v), unless aged P16, when the procedure was performed terminally.

Generation of *Ikzf2del890* mice. The *Ikzf2del890* mutant line was generated by the Molecular and Cellular Biology group at the Mary Lyon Centre, MRC Harwell Institute using CRISPR–Cas9 gene editing, as described previously[37] (see Supplementary Table 13 for single-guide RNA (sgRNA) sequences, donor oligonucleotide sequences and genotyping primers). For construction of each sgRNA plasmid, a pair of single-stranded donor oligonucleotides (IDT) was hybridized and cloned using Gibson Assembly Master Mix (NEB) into linearized p_1.1 plasmid digested with StuI and AflII to express sgRNAs under the T7 promoter.

The p_1.1_sgRNA plasmids were linearized with XbaI, purified with phenol–chloroform, and the products were used as templates from which sgRNAs were in vitro transcribed. sgRNAs were synthesized using MEGAshortscript T7 Transcription Kit (Ambion). RNAs were purified using MEGAclear Transcription Clean-Up Kit (Ambion). RNA quality was assessed using a NanoDrop (Thermo Scientific) and by electrophoresis on 2% agarose gel containing Ethidium Bromide (Fisher Scientific).

As this exon deletion mutant was generated as part of an experiment to generate a floxed mutant, a *Ikzf2* flox long single-stranded DNA (lssDNA) donor was also synthesized as described previously for inclusion in the microinjection mix[38].

For microinjections, the pronucleus of one-cell stage C57BL/6NTac embryos were injected with a mix containing *Cas9* mRNA (5meC,Ψ, Tebu-Bio/TriLink Biotechnologies) at 100 ng μl^{-1}, the four *Ikzf2* sgRNAs, each at 50 ng μl^{-1} and the *Ikzf2* flox lssDNA donor at 50 ng μl^{-1} prepared in microinjection buffer. Injected embryos were re-implanted in pseudo-pregnant CD-1 females, which were allowed to litter and rear F_0 progeny.

For genotyping, genomic DNA was extracted from ear biopsies of F_0 and F_1 mice using DNA Extract All Reagents Kit (Applied Biosystems) and amplified by PCR using high fidelity Expand Long Range dNTPack (Roche) and specific genotyping

primers (see Supplementary Table 13). PCR products were further purified using QIAquick Gel Extraction Kit (Qiagen) and analysed by Sanger sequencing. Copy counting experiments by droplet digital PCR (ddPCR) against a known two copy reference (*Dot1l*) were also carried out to confirm the exon deletion and that there were no additional integrations of the lssDNA donor. Mice carrying the *del890* deletion allele were subsequently mated with mice carrying the *cello* mutation to generate *Ikzf2*^cello/del890 compound heterozygotes for complementation testing.

Scanning electron microscopy. Mice were euthanized by cervical dislocation and inner ears were removed and fixed in 2.5% glutaraldehyde (TAAB Laboratories Equipment Ltd) in 0.1 M phosphate buffer for 4 h at 4 °C. After decalcification in 4.3% EDTA, cochleae were dissected to expose the organ of Corti, and subjected to 'OTO' processing (1 h incubation in 1% osmium tetroxide (TAAB Laboratories Equipment), 30 min incubation in 1% thiocarbohydrazide (Sigma), 1 h incubation in 1% osmium tetroxide), before dehydration in increasing concentrations of ethanol (25%, 40%, 60%, 80%, 95%, 2 × 100%) at 4 °C. Samples were critical point dried with liquid CO_2 using an Emitech K850 (EM Technologies), then mounted on stubs using silver paint (Agar Scientific) and sputter coated with platinum using a Quorum Q150R S sputter coater (Quorum Technologies). Samples were examined using a JEOL JSM-6010LV Scanning Electron Microscope. Hair cell bundle counts were performed by counting the number of OHC and IHC bundles adjacent to ten pillar cells in the apical (<180° from apex), mid (180–450° from apex) and basal (>450° from apex) regions of the cochlea. At least three ears (one ear per mouse) were analysed for each genotype at each time point.

Electrophysiological analyses. Electrophysiological recordings were made from OHCs of *cello* mice aged P9–P18. Cochleae were dissected in normal extracellular solution (in mM): 135 NaCl, 5.8 KCl, 1.3 CaCl$_2$, 0.9 MgCl$_2$, 0.7 NaH$_2$PO$_4$, 5.6 D-glucose, 10 HEPES-NaOH. Sodium pyruvate (2 mM), MEM amino acids solution (50×, without L-glutamine) and MEM vitamins solution (100×) were added from concentrates (Fisher Scientific). The pH was adjusted to 7.5 (osmolality approximately 308 mmol kg^{-1}). The dissected cochleae were transferred to a microscope chamber, immobilized as previously described[39] and continuously perfused with a peristaltic pump using the above extracellular solution. The organs of Corti were viewed using an upright microscope (Nikon FN1) with Nomarski optics (60× objective).

MET currents were elicited by stimulating the hair bundles of P9 OHCs in the excitatory and inhibitory direction using a fluid jet from a pipette (tip diameter 8–10 μm) driven by a piezoelectric disc[39]. The pipette tip of the fluid jet was positioned near to the bundles to elicit a maximal MET current. Mechanical stimuli were applied as 50 Hz sinusoids (filtered at 0.25 kHz, 8-pole Bessel) with driving voltages of ±40 V. MET currents were recorded with a patch pipette solution containing (in mM): 106 Cs-glutamate, 20 CsCl, 3 MgCl$_2$, 1 EGTA-CsOH, 5 Na$_2$ATP, 0.3 Na$_2$GTP, 5 HEPES-CsOH, 10 sodium phosphocreatine (pH 7.3). Membrane potentials were corrected for the liquid junction potential (–11 mV).

Patch clamp recordings were performed using an Optopatch (Cairn Research) amplifier. Patch pipettes were made from soda glass capillaries (Harvard Apparatus) and had a typical resistance in extracellular solution of 2–3 MΩ. To reduce the electrode capacitance, patch electrodes were coated with surf wax (Mr. Zog's SexWax). Potassium current recordings were performed at room temperature (22–24 °C) and the intracellular solution contained (in mM): 131 KCl, 3 MgCl$_2$, 1 EGTA-KOH, 5 Na$_2$ATP, 5 HEPES-KOH, 10 Na$_2$-phosphocreatine (pH 7.3; osmolality approximately 296 mmol kg^{-1}). Data acquisition was controlled by pClamp software (version 10) using Digidata 1440A boards (Molecular Devices). Recordings were low-pass filtered at 2.5 kHz (8-pole Bessel), sampled at 5 kHz and stored on computer for off-line analysis (Origin, OriginLab). Membrane potentials in voltage clamp were corrected for the voltage drop across the uncompensated residual series resistance and for a liquid junction potential (−4 mV).

The presence of electromotile activity in P16–P18 OHCs was estimated by applying a depolarizing voltage step from the holding potential of −64 mV to +56 mV. Changes in cell length were viewed and recorded with a Nikon FN1 microscope (75× magnification) with a Flash 4.0 SCCD camera (Hamamatsu). Cell body movement was tracked using Fiji software. Lines were drawn across the basal membrane of patched OHCs, perpendicular to the direction of cell motion, and a projected time-based z-stack of the pixels under the line was made. Cell movement was measured with Photoshop as a pixel shift and then converted to nanometres (290 pixels = 10 μm).

Nonlinear (voltage-dependent) capacitance of IHCs in Anc80-*Ikzf2*-injected mice and their non-injected littermates was studied at P12–P16 using conventional whole-cell patch clamp recordings. Apical turn of the organ of Corti was carefully dissected in Leibovitz's L-15 cell culture medium (21083027, Gibco/ThermoFisher) containing the following inorganic salts (in mM): 137 NaCl, 5.4 KCl, 1.26 CaCl$_2$, 1.0 MgCl$_2$, 1.0 Na$_2$HPO$_4$, 0.44 KH$_2$PO$_4$ and 0.81 MgSO$_4$ and placed into a custom-made recording chamber, where it was held by two strands of dental floss. The organ of Corti explants were viewed with an upright microscope (BX51WIF, Olympus), equipped with a high numerical aperture (NA) objective (100×, 1.0

NA). To block voltage-gated ion channels in IHCs, the bath solution was made of L-15 medium supplemented with 10 mM tetraethylammonium-Cl, 2 mM CoCl$_2$, 10 mM CsCl and 0.1 mM nifedipine (all from Sigma), while the intrapipette solution contained (in mM): 140 CsCl, 2.5 MgCl$_2$, 2.5 Na$_2$ATP, 1.0 EGTA and 5 HEPES. During recordings, the organs of Corti were continuously perfused with the above extracellular bath solution. Whole-cell current responses were recorded with MultiClamp 700B patch clamp amplifier (Molecular Devices), controlled by jClamp software (SciSoft). Membrane capacitance was measured during the voltage ramp with a dual sinusoidal, FFT-based method[40]. The recorded capacitance was fitted to the first derivative of a two-state Boltzmann function that is typically used to fit nonlinear capacitance of OHCs plus a small correction for the membrane area changes between expanded and contracted states of prestin[41], as follows:

$C_m = C_v + C_{lin}$, in which C_m is the total membrane capacitance, C_v is a voltage-dependent (nonlinear) component, and C_{lin} is a voltage-independent (linear) component.

$$ C_V = Q_{max} \frac{ze}{kT} \frac{b}{(1+b)^2} + \frac{\Delta C_{sa}}{(1+b^{-1})} ; b = \exp\left(\frac{-ze(V-V_{pk})}{kT} \right) $$

in which Q_{max} is the maximum nonlinear charge moved, V_{pk} is a voltage at peak capacitance, V is membrane potential, z is valence, e is electron charge, k is Boltzmann's constant, T is absolute temperature, and ΔC_{sa} is the maximum increase in capacitance that occurs when all prestin molecules change from compact to expanded state. To account for some variability in sizes of IHCs, statistical data are shown as the maximum of voltage-dependent component of capacitance (C_v) normalized to the linear capacitance of the cell (C_v/C_{lin}).

DPOAEs. DPOAE tests were performed using frequency-specific tone-burst stimuli at 8, 16 and 32 kHz with the TDT RZ6 System 3 hardware and BioSig RZ (version 5.7.1) software (Tucker Davis Technology). An ER10B+ low noise probe microphone (Etymotic Research) was used to measure the DPOAE near the tympanic membrane. Tone stimuli were presented via separate MF1 (Tucker Davis Technology) speakers, with f_1 and f_2 at a ratio of $f_2/f_1 = 1.2$ (L1 = 65 dB SPL, L2 = 55 dB SPL), centred around the frequencies of 8, 16 and 32 kHz. Surgical anaesthesia was achieved by intraperitoneal injection of ketamine (100 mg ml^{-1} at 10% v/v), xylazine (20 mg ml^{-1} at 5% v/v) and acepromazine (2 mg ml^{-1} at 8% v/v) administered at a rate of 0.1 ml per 10 g body mass. Once the required depth of anaesthesia was confirmed by the lack of the pedal reflex, a section of pinna was removed to allow unobstructed access to the external auditory meatus. Mice were then placed on a heated mat inside a sound-attenuated chamber (ETS-Lindgren) and the DPOAE probe assembly was inserted into the ear canal using a pipette tip to aid correct placement. In-ear calibration was performed before each test. The f_1 and f_2 tones were presented continuously and a fast-Fourier transform was performed on the averaged response of 356 epochs (each approximately 21 ms). The level of the $2f_1 - f_2$ DPOAE response was recorded and the noise floor calculated by averaging the four frequency bins either side of the $2f_1 - f_2$ frequency.

NanoString validation. Cochlear RNA extracted from biological triplicates of *Ikzf2*^cello/cello, *Ikzf2*^cello/+ and *Ikzf2*^+/+ animals at P8 were processed for NanoString validation at the UMSOM Institute for Genome Sciences using the nCounter Master Kit per manufacturer's instructions, and quantified using the NanoString nCounter platform. See Supplementary Table 13 for NanoString probe sequences. Data were analysed using nSolver 4.0 software (NanoString).

Anc80L65 AAV vector construction. The Anc80L65-Myc-*Ikzf2*+ (Anc80-*Ikzf2*) expression vector was designed to drive expression of a Myc-tagged *Ikzf2* construct followed by a bovine Growth Hormone poly-adenylation (BGH pA) site under control of the cytomegalovirus (CMV) promoter. The Anc80L65-eGFP (Anc80-eGFP) expression construct also contained a Woodchuck Hepatatis Virus Posttranscriptional Regulatory Element (WPRE) preceding the BGH pA site. Anc80L65 AAV vectors[17,18] were produced by the Gene Transfer Vector Core, Grousbeck Gene Therapy Center at the Massachusetts Eye and Ear Infirmary (http://vector.meei.harvard.edu/).

Inner ear gene delivery. For in vivo hair cell transductions, mice were injected with Anc80L65 AAVs between P1 and P3 via the posterior semicircular canal using the injection method described previously[42]. In brief, animals were anesthetized on ice before a post-auricular incision was made on either the left or right side. Tissues were further dissected to reveal the posterior semicircular canal, and a Nanolitre 2010 microinjection system (World Precision Instruments) equipped with a loaded glass needle was used to inject 700 nl of 1.13×10^{13} genome copies (GC) per ml Anc80-*Ikzf2* or 500 nl of 4.85×10^{12} GC per ml Anc80-eGFP. Injections into the inner ear were performed in 50 nl increments over the course of 2 min. The needle was then removed, the incision sutured, and animals were placed on a 37 °C heating pad to recover before being returned to their cage.

FACS. For the scRNA-seq analysis of Anc80-*Ikzf2* transduced hair cells, inner ears of neonatal *Myo15*^cre/+;*ROSA26*^CAG-tdTomato mice were injected with Anc80-*Ikzf2* (4 mice) or control Anc80-eGFP (2 mice) via the posterior semicircular canal.

Cochlear tissues from both injected and uninjected ears were obtained at P8 and further dissected to reveal the sensory epithelium. Inclusion of the uninjected ear in the single cell analysis allowed for the study of changes in gene expression that occur in response to a gradient of transgene expression. This is because, in mice, inner ear gene delivery often results in transduction in the contralateral ear, albeit at a lower intensity[18]. Cochlear tissues were then dissociated for fluorescence activated cell sorting (FACS) using the method described previously[9]. In brief, the sensory epithelia from Anc80-eGFP- and Anc80-*Ikzf2*-injected mice were pooled separately into two wells of a 48-well plate containing 0.5 mg ml^{-1} thermolysin (Sigma). Tissues were incubated at 37 °C for 20 min, after which the thermolysin was removed and replaced with accutase enzyme (MilliporeSigma). After a 3-min incubation at 37 °C, tissues were mechanically disrupted using a 23G blunt ended needle connected to a 1 ml syringe. This step was performed twice. After confirming tissue dissociation by direct visualization, the dissociation reaction was stopped by adding an equal volume of IMDM supplemented with 10% heat-inactivated FBS to the Accutase enzyme solution. Cells were passed through a 40 mm cell strainer (BD) to remove cell clumps. tdTomato-expressing hair cells were sorted in ice-cold tubes containing IMDM with 10% FBS on a BD FACSAria II (BD Biosciences) and processed for scRNA-seq. Flow cytometry analyses were performed with assistance from X. Fan at the University of Maryland Marlene and Stewart Greenebaum Comprehensive Cancer Center Flow Cytometry Shared Service.

scRNA-seq. tdTomato-positive sorted hair cells were pelleted once (300*g* at 4 °C) and resuspended in a minimal remaining volume (around 30 μl). Hair cell-enriched single-cell suspensions were then used as input on the 10× Genomics Chromium platform with 3′ Single Cell v2 chemistry (10× Genomics). After capture and library preparation, scRNA-seq libraries were sequenced on a NextSeq 500 (Illumina) in collaboration with the NIDCD Genomics and Computational Biology Core. Samples were sequenced to an average depth of over 300,000 reads per cell, which resulted in detection of a median of >3,000 genes (Anc80-eGFP) and >4,000 genes (Anc80-*Ikzf2*) per cell, ensuring maximal transcriptional complexity and detection of low-abundance transcripts (see Extended Data Fig. 9b, c). Reads were aligned to a modified mm10 mouse reference containing the sequences for the Ai14 locus, as well as Anc80-eGFP and Anc80-*Ikzf2* viral sequences (Extended Data Fig. 9a) using the 10× Genomics Cell Ranger (version 2.0.2) package to generate the read counts matrix files. Read counts from viral and Ai14 loci were removed from the expression matrix before dimensionality reduction so as to not influence data clustering. Cells from these hair cell clusters were determined to be Anc80-*Ikzf2*(+) versus Anc80-*Ikzf2*(−), and IHCs versus OHCs, based on their expression of Anc80-*Ikzf2* and *Slc17a8*, respectively (Fig. 3, Extended Data Figs. 8 and 9, Supplementary Table 9). *Slc26a5* was not well detected in the scRNA-seq dataset and was therefore not used as an OHC marker. After clustering, four hair cells were excluded based on co-expression of a contaminating cell type. Secondary analyses, including shared nearest neighbour (SNN) clustering, *t*-SNE embedding, and differential expression testing (using either Wilcoxon ranked sum for marker gene identification or MAST for pairwise comparison between control IHCs and OHCs) were performed in R with Seurat (version 2.1.0)[43,44]. Non-parametric analysis of variance between the four classified groups of HCs (IHCs and OHCs with either high or low Anc80-*Ikzf2* expression) using a Kruskal–Wallis test was performed to help qualify genes that had statistical difference across these cell populations. This was followed by post hoc pairwise Wilcoxon ranked sum comparisons to assess multiple-comparison-adjusted *P* values. Additional plots were generated by NMF (version 0.20.6) and ggplot2 (version 2.2.1)[45,46]. These analyses used the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).

Immunohistochemistry of AAV-injected cochleae. Mouse inner ears injected with either Anc80-*Ikzf2* or Anc80-eGFP were between P8 and 8 weeks, fixed in 4% PFA in PBS overnight at 4 °C, and decalcified in a solution of 5% EDTA in RNAlater (Invitrogen). Decalcified ears were processed by sucrose gradient and embedded in OCT compound (Tissue-Tek) for cryosectioning, or fine dissected for whole-mount immunohistochemistry. Cryosections (10 μm) on positively charged glass slides were used for in situ hybridization (ISH) and section immunohistochemistry. For whole-mount immunolabelling at 6–8 weeks, hair cell loss was observed in the injected ear and therefore the contralateral ear, expressing a lower level of the Anc80-*Ikzf2* virus, was used. Primary antibodies: goat anti-prestin N-20 (1:200, Santa Cruz Biotechnology); goat anti-oncomodulin N-19 (1:100, Santa Cruz Biotechnology); rabbit anti-myosinVI (1:1,000, Proteus BioSciences); rabbit anti-GFP (1:100, Life Technologies); mouse anti-cMyc 9E10 (1:100, Santa Cruz Biotechnology) and mouse anti-otoferlin (1:100, Abcam). The guinea pig anti-VGLUT3 antibody (1:5,000) used in this study was donated by R. Seal. Corresponding Alexa Fluor 488 and 546 (1:800, Invitrogen) were used for secondary detection, Alexa Fluor 488 Phalloidin (1:1,000, Invitrogen) was used to mark F-actin, and DAPI (1:20,000, Thermo Fisher) was used to mark cell nuclei. Images were acquired using a Nikon Eclipse E600 microscope (Nikon) equipped with a Lumenera Infinity 3 camera. Whole-mount images were acquired using a Zeiss

LSM DUO confocal microscope, located at the UMSOM Confocal Microscopy Core, at 63× oil magnification. Images were processed using Infinity Capture and Infinity Analyze software (Lumenera), and ImageJ software.

RNA in situ hybridization. In situ hybridization was performed as described previously[47]. In brief, slides were re-fixed in 4% PFA, and then treated with 2 μg ml^{-1} Proteinase-K for 10 min. Proteinase K reaction was stopped by soaking slides again in 4% PFA, followed by acetylation and permeabilization. Hybridization for the digoxigenin labelled *Fcrlb* probe was performed overnight at 65 °C (see Supplementary Table 13 for *Fcrlb* probe primers). After a series of washes in saline sodium citrate, slides were incubated with sheep-anti-digoxigenin antibody conjugated to alkaline phosphatase (Sigma-Aldrich, 1:100) overnight at 4 °C. Slides were then incubated in BM purple AP substrate precipitating solution (Roche) to localize bound anti-digoxigenin antibody.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The RiboTag OHC RNA-seq, P8 cello cochlea RNA-seq, and P8 Anc80-*Ikzf2* and Anc80-eGFP injected cochlea scRNA-seq data have been submitted to the Gene Expression Omnibus (GEO) database under accessions GSE116703, GSE116702 and GSE120462, and are also available for viewing through the gEAR portal (https://umgear.org/).

23. Caberlotto, E. et al. Usher type 1G protein sans is a critical component of the tip-link complex, a structure controlling actin polymerization in stereocilia. *Proc. Natl Acad. Sci. USA* **108**, 5825–5830 (2011).
24. Johnson, K. R., Zheng, Q. Y. & Noben-Trauth, K. Strain background effects and genetic modifiers of hearing in mice. *Brain Res.* **1091**, 79–88 (2006).
25. Hoelter, S. M. et al. "Sighted C3H" mice—a tool for analysing the influence of vision on mouse behaviour? *Front. Biosci.* **13**, 5810–5823 (2008).
26. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
27. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
28. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
29. Ulitsky, I. et al. Expander: from expression microarrays to networks and functions. *Nat. Protocols* **5**, 303–322 (2010).
30. Sharan, R., Maron-Katz, A. & Shamir, R. CLICK and EXPANDER: a system for clustering and visualizing gene expression data. *Bioinformatics* **19**, 1787–1799 (2003).
31. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
32. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
33. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688 (2012).
34. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protocols* **4**, 1073–1081 (2009).
35. Källberg, M. et al. Template-based protein structure modeling using the RaptorX web server. *Nat. Protocols* **7**, 1511–1522 (2012).
36. Hardisty-Hughes, R. E., Parker, A. & Brown, S. D. M. A hearing and vestibular phenotyping pipeline to identify mouse mutants with hearing impairment. *Nat. Protocols* **5**, 177–190 (2010).
37. Mianné, J. et al. Correction of the auditory phenotype in C57BL/6N mice via CRISPR/Cas9-mediated homology directed repair. *Genome Med.* **8**, 16 (2016).
38. Codner, G. F. et al. Application of long single-stranded DNA donors in genome editing: generation and validation of mouse mutants. *BMC Biol.* **16**, 70 (2018).
39. Corns, L. F., Johnson, S. L., Kros, C. J. & Marcotti, W. Calcium entry into stereocilia drives adaptation of the mechanoelectrical transducer current of mammalian cochlear hair cells. *Proc. Natl Acad. Sci. USA* **111**, 14918–14923 (2014).
40. Santos-Sacchi, J. Determination of cell capacitance using the exact empirical solution of partial δY/δC$_m$ and its phase angle. *Biophys. J.* **87**, 714–727 (2004).
41. Santos-Sacchi, J. & Navarrete, E. Voltage-dependent changes in specific membrane capacitance caused by prestin, the outer hair cell lateral membrane motor. *Pflugers Arch.* **444**, 99–106 (2002).
42. Isgrig, K. et al. Gene therapy restores balance and auditory functions in a mouse model of usher syndrome. *Mol. Ther.* **25**, 780–791 (2017).
43. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
44. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
45. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2009).
46. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
47. Geng, R. et al. Comprehensive expression of Wnt signaling pathway genes during development and maturation of the mouse cochlea. *PLoS One* **11**, e0148339 (2016).

a *Prestin^{CreERT2/+};ROSA26^{CAG-tdTomato}*

c

Cell type	Gene	Prestin^{CreERT2}; RiboTag IP				
		Log$_2$EF P8	Log$_2$EF P14	Log$_2$EF P28	Log$_2$EF 6wk	Log$_2$EF 10wk
Outer hair cells	*Ocm*	3.42	2.75	3.2	2.69	4.95
	Slc26a5	2.68	0.57	-0.06	0.47	2.91
	Strc	3.04	1.08	2.69	1.17	2.74
All hair cells	*Gfi1*	2.25	1.32	0.21	1.04	2.33
	Pou4f3	3.83	0.91	0.5	1.09	2.58
Inner hair cells	*Atp2a3*	-0.65	-2.01	-1.31	-1.72	-2.02
	Otof	0.46	-1.16	-1.03	-1.12	-0.25
	Slc17a8	-0.27	-0.68	-0.86	-1.77	-1.35
Supporting cells	*Sox2*	0.28	-0.03	0.71	0.29	-0.48
	Jag1	-0.6	-1.11	-1.00	-1.11	-0.24
Neurons	*Tubb3*	-1.91	-1.2	-0.02	-0.59	-1.13
Mesenchyme	*Pou3f4*	-0.08	-0.61	-1.17	-0.82	-1.11

b *RiboTag^{HA/+};Prestin^{CreERT2/+}*

d

e

f

Cluster	GO category	*p*-value	Empirical *p*-value
Cluster 1	anatomical structure morphogenesis - GO:0009653	3.51E-08	0.001
	skeletal system morphogenesis - GO:0048705	2.17E-06	0.003
	locomotion - GO:0040011	3.98E-06	0.006
	extracellular matrix structural constituent - GO:0005201	8.34E-06	0.018
	cell motility - GO:0048870	3.53E-05	0.047
Cluster 2	cell development - GO:0048468	3.52E-11	0.001
	voltage-gated cation channel activity - GO:0022843	3.08E-06	0.005
	transmission of nerve impulse - GO:0019226	1.07E-05	0.02

Extended Data Fig. 1 | RiboTag immunoprecipitation enriches for known OHC-expressed transcripts. a, Representative *prestin^{creERT2/+};ROSA26^{CAG-tdTomato}* cochlear whole-mount. The *prestin^{creERT2}*-driven tdTomato expression is OHC-specific at P21 ($n = 1$). Scale bar, 20 μm. **b**, Schematic of the RiboTag immunoprecipitation protocol. Red OHCs represent Cre/HA-tagged ribosome expression. **c**, RiboTag RNA-seq log$_2$ enrichment and depletion of transcripts for known inner ear cell type markers (EF = log$_2$(IP/input)). **d**, Genes at least two-fold enriched in IHCs ($n = 565$ genes) or OHCs ($n = 253$ genes) in the published dataset[8] are significantly depleted or enriched, respectively, by the RiboTag OHC immunoprecipitation at all time points examined

(two-sided Wilcoxon's test). Black line represents median EF, box demarcates first and third quartiles, whiskers demarcate first and third quartiles $\pm 1.5 \times$ IQR values, dots represent single outliers. **e**, Clustering of genes differentially expressed across OHC postnatal development. Error bars denote s.d. Before clustering, expression levels were standardized to mean $= 0$ and s.d. $= 1$. **f**, Enriched Gene Ontology (GO) functional categories identified for the gene clusters in **e** (cluster 1 $n = 160$ genes, cluster 2 $n = 63$ genes). No significantly enriched GO categories were found for cluster 3 ($n = 79$ genes). Enrichment and statistical analyses were performed using the EXPANDER implemented tool TANGO.
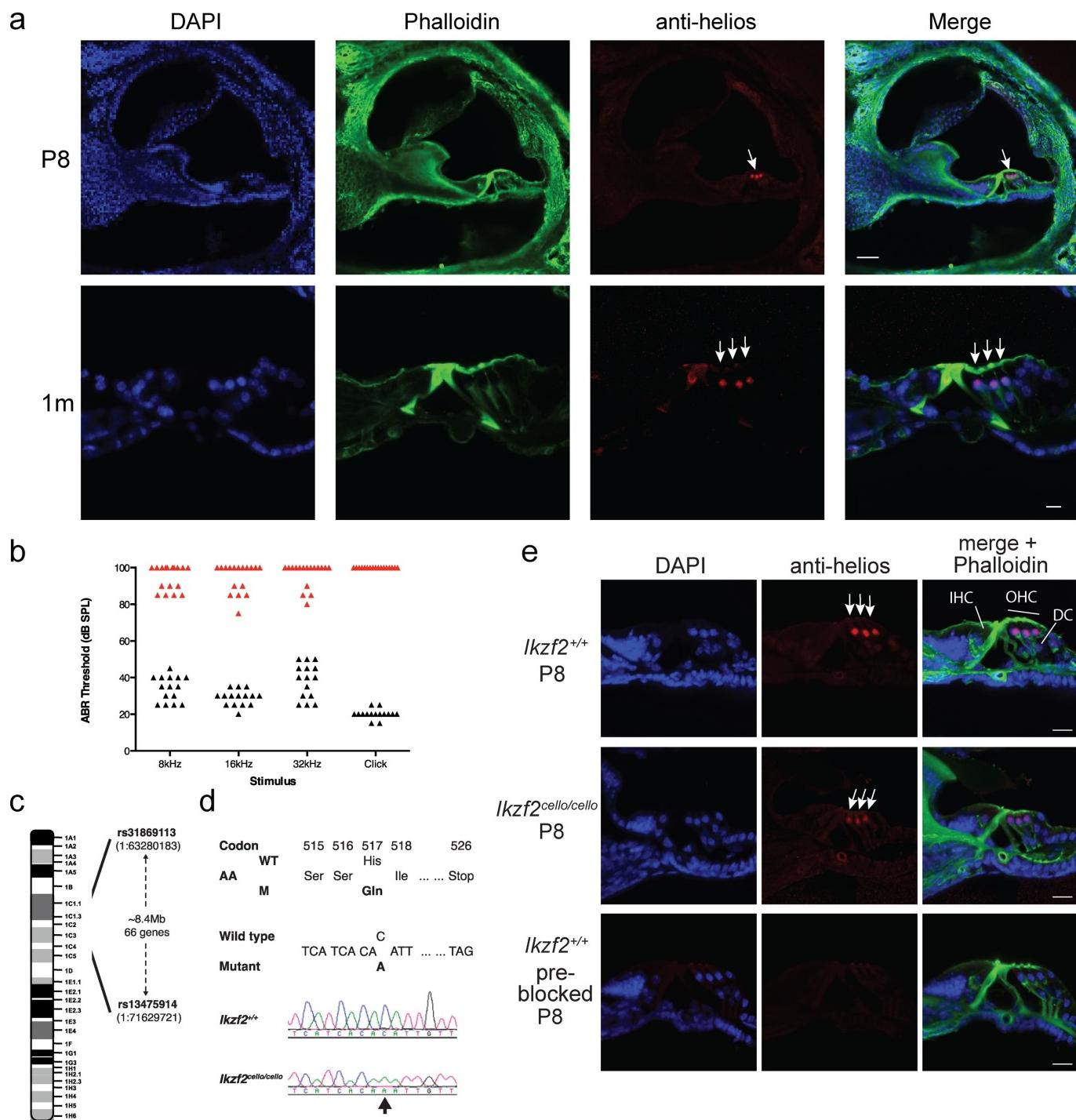
Extended Data Fig. 2 | Auditory phenotyping, SNP mapping and whole-genome sequencing of mouse pedigree MPC173, subsequently named *cello*. **a**, Specific expression of Helios can be seen in the nuclei of wild-type P8 OHCs (white arrow), and is maintained in wild-type OHCs at 1 month (white arrows). Scale bars, 50 μm (P8) and 10 μm (1 month). $n = 3$ biologically independent samples for each time point. **b**, Auditory brainstem response phenotyping of pedigree MPC173 at 9 months of age identified 17 biologically independent animals with increased hearing thresholds (red triangles) compared to their normal hearing colony mates (black triangles). $n = 15$ biologically independent animals. **c**, The mutation mapped to an 8.4-Mb region on chromosome 1 between single nucleotide polymorphism (SNP) rs31869113 and

rs13475914 (chr1: 63280183–71629721), containing 66 genes. **d**, Detection of a non-synonymous mutation in *cello*. DNA sequencing identified a nucleotide transversion (c.1551C>A) in the *Ikzf2* gene at codon 517, thus altering the wild-type sequence CAC, encoding histidine, to the mutant (M) sequence CAA, encoding glutamine. Electropherograms derived from a *cello* mutant mouse (*Ikzf2*$^{cello/cello}$) and a wild-type colony mate (*Ikzf2*$^{+/+}$) control showing the sequence surrounding *Ikzf2* nucleotide 1551 (indicated by an arrow). **e**, Helios is expressed in the OHC nuclei of both *Ikzf2*$^{+/+}$ and *Ikzf2*$^{cello/cello}$ mice at P8. $n = 3$ biologically independent samples per genotype. Loss of labelling when the anti-Helios antibody is 'pre-blocked' confirms specificity. $n = 1$ biologically independent sample. Scale bars, 20 μm. DC, Deiters' cells.
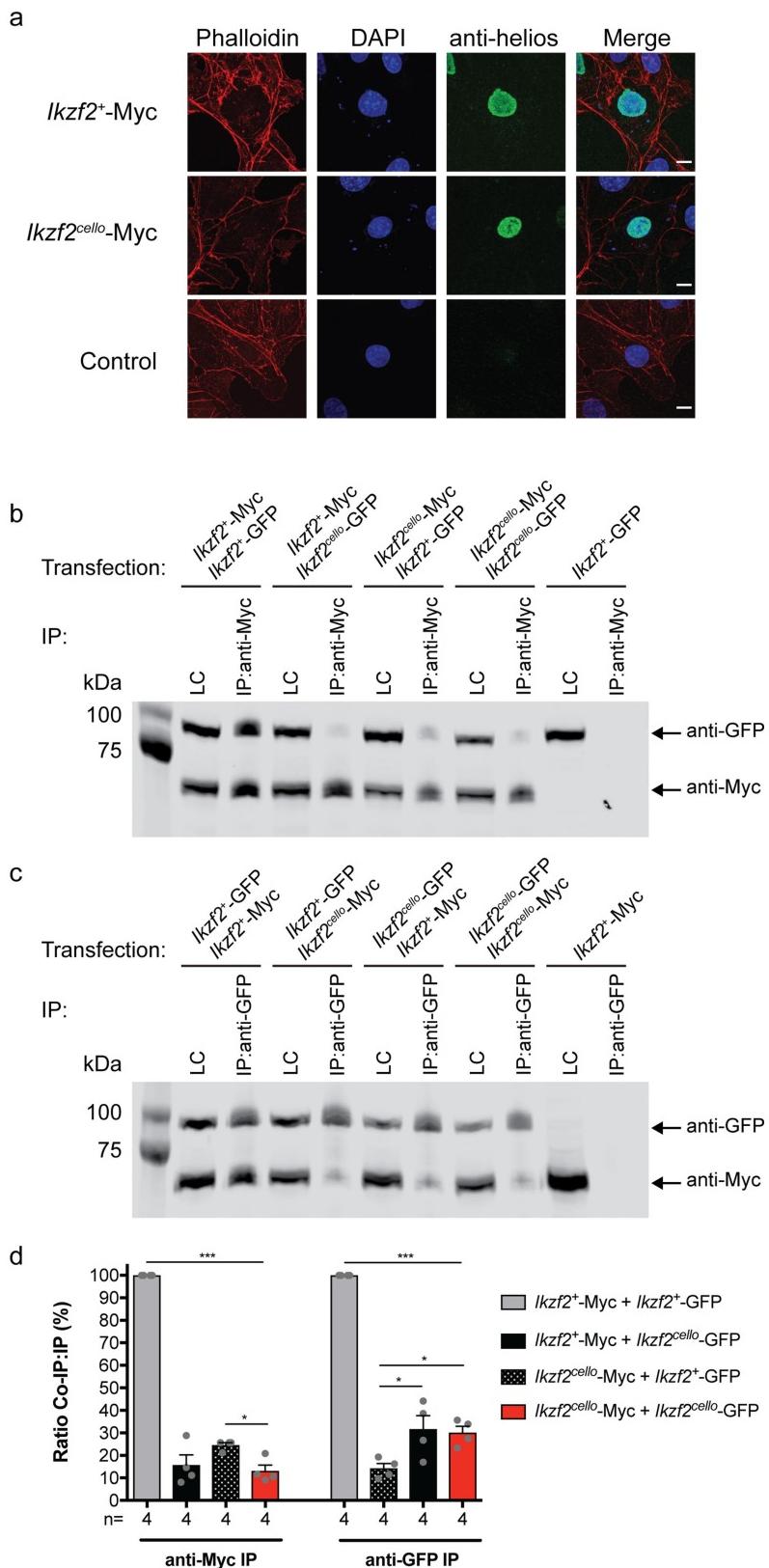
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | The *Ikzf2^cello* mutation disrupts homodimerization of Helios. a, Cos-7 cells transfected with *Ikzf2^+*- or *Ikzf2^cello*-Myc. Nuclear localization is unaffected by the *Ikzf2^cello* mutation. $n = 2$ biologically independent experiments. Scale bars, 10 μm. **b**, Co-immunoprecipitation (IP) of Myc-tagged (62 kDa) and GFP-tagged (88 kDa) *Ikzf2^+* and *Ikzf2^cello* constructs. Transfected cell lysates were immunoprecipitated using an anti-Myc antibody and analysed by western blotting with both anti-Myc and anti-GFP antibodies. Results show that wild-type *Ikzf2^+* Helios can dimerize, but that dimerization is impaired by the *cello* mutation. LC, cell lysate loading control. **c**, Reciprocal immunoprecipitation reactions using an anti-GFP antibody confirm dimerization of wild-type *Ikzf2^+* Helios and reduced dimerization of mutant *Ikzf2^cello* Helios. **d**, Quantification of co-immunoprecipitation western blots. Band intensities were determined and used to calculate the relative ratio of the co-immunoprecipitation to immunoprecipitation signal. $n = 4$ biologically independent experiments. Data are mean ± s.e.m. Anti-Myc IP: ***$P < 0.0001$ (*Ikzf2^+*-Myc + *Ikzf2^+*-GFP vs *Ikzf2^+*-Myc + *Ikzf2^cello*-GFP, vs *Ikzf2^cello*-Myc + *Ikzf2^+*-GFP and vs *Ikzf2^cello*-Myc + *Ikzf2^cello*-GFP). *$P = 0.0476$ (*Ikzf2^cello*-Myc + *Ikzf2^+*-GFP vs *Ikzf2^cello*-Myc + *Ikzf2^cello*-GFP). $P = 0.1488$ (*Ikzf2^+*-Myc + *Ikzf2^cello*-GFP vs *Ikzf2^cello*-Myc + *Ikzf2^+*-GFP). $P = 0.9020$ (*Ikzf2^+*-Myc + *Ikzf2^cello*-GFP vs *Ikzf2^cello*-Myc + *Ikzf2^cello*-GFP). Anti-GFP IP: ***$P < 0.0001$ (*Ikzf2^+*-Myc + *Ikzf2^+*-GFP vs *Ikzf2^cello*-Myc + *Ikzf2^+*-GFP, vs *Ikzf2^+*-Myc + *Ikzf2^cello*-GFP and vs *Ikzf2^cello*-Myc + *Ikzf2^cello*-GFP). *$P = 0.0202$ (*Ikzf2^cello*-Myc + *Ikzf2^+*-GFP vs *Ikzf2^+*-Myc + *Ikzf2^cello*-GFP) *$P = 0.0346$ (*Ikzf2^cello*-Myc + *Ikzf2^+*-GFP vs *Ikzf2^cello*-Myc + *Ikzf2^cello*-GFP). $P = 0.9894$ (*Ikzf2^+*-Myc + *Ikzf2^cello*-GFP vs + *Ikzf2^cello*-Myc + *Ikzf2^cello*-GFP) (one-way ANOVA with Tukey post hoc test). See Supplementary Fig. 1 for source images.
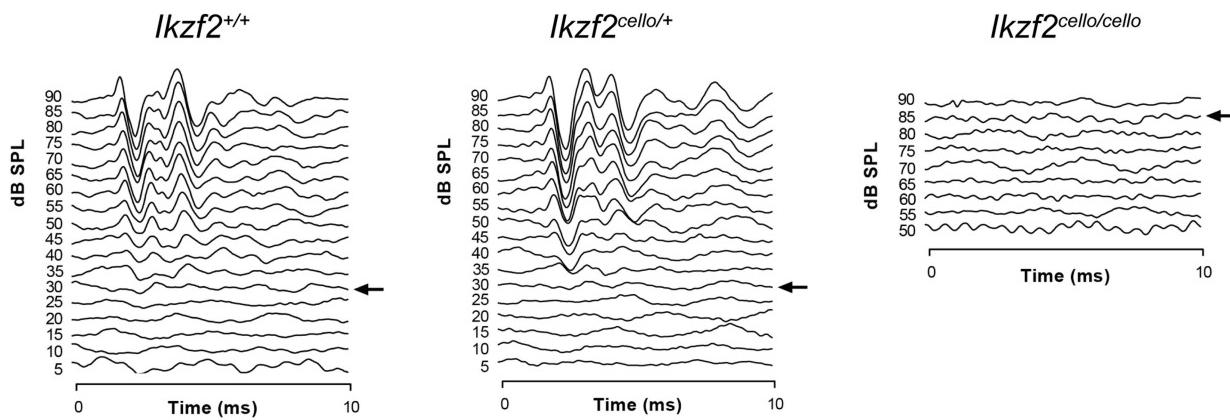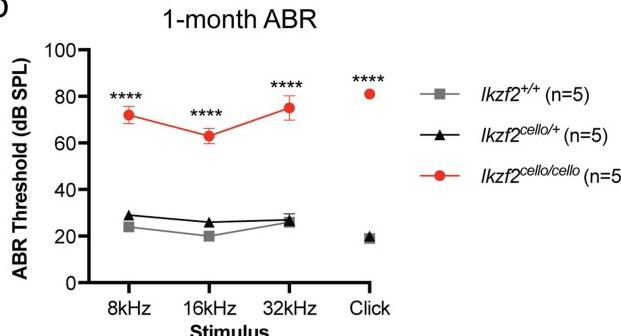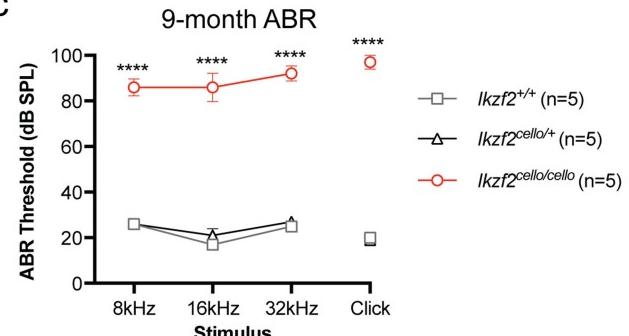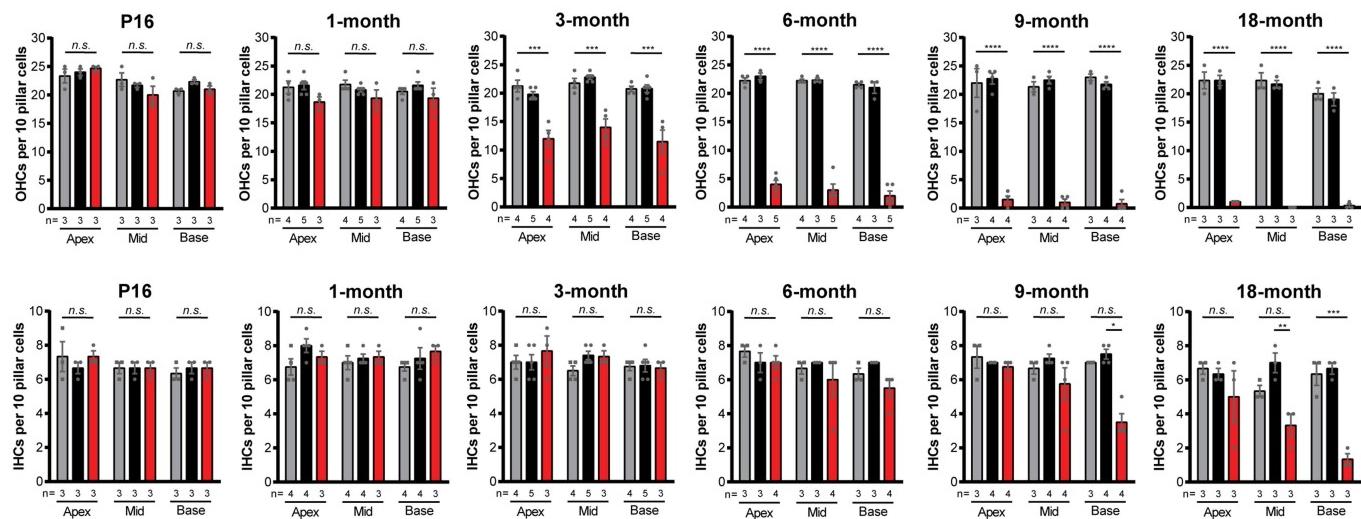
Extended Data Fig. 4 | Auditory function and HC bundle survival in cello mice. a, Representative click ABR waveforms for *Ikzf2*[+/+], *Ikzf2*[cello/+] and *Ikzf2*[cello/cello] littermates at P16. $n = 4$ biologically independent animals per genotype. **b, c**, Averaged ABR thresholds for *cello* mice at 1-month of age (**b**) and 9 months of age (**c**). Age-matched *Ikzf2*[+/+] and *Ikzf2*[cello/+] controls display thresholds within the expected range (15–30 dB SPL) at all time points tested. $n = 5$ biologically independent animals per genotype. Data are mean thresholds $\pm$ s.e.m. 1-month *Ikzf2*[cello/cello] vs *Ikzf2*[+/+]: ****$P < 0.0001$ (8 kHz, 16 kHz, 32 kHz, click). 1-month *Ikzf2*[cello/cello] vs 1 *Ikzf2*[cello/+]: ****$P < 0.0001$ (8 kHz, 16 kHz, 32 kHz, click). 9-month

Ikzf2[cello/cello] vs *Ikzf2*[+/+]: ****$P < 0.0001$ (8 kHz, 16 kHz, 32 kHz, click). 9-month *Ikzf2*[cello/cello] vs *Ikzf2*[cello/+]: ****$P < 0.0001$ (8 kHz, 16 kHz, 32 kHz, click) (one-way ANOVA with Tukey post hoc test). **d**, OHC and IHC bundle counts for *cello* mice from P16 to 18 months of age. Grey, *Ikzf2*[+/+]; black, *Ikzf2*[cello/+]; red, *Ikzf2*[cello/cello]. Data are mean $\pm$ s.e.m. *n.s.*, non-significant. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ (one-way ANOVA with Tukey post hoc test). Number of biologically independent samples for OHC and IHC bundle counts are shown. See also Supplementary Table 5 and 6.
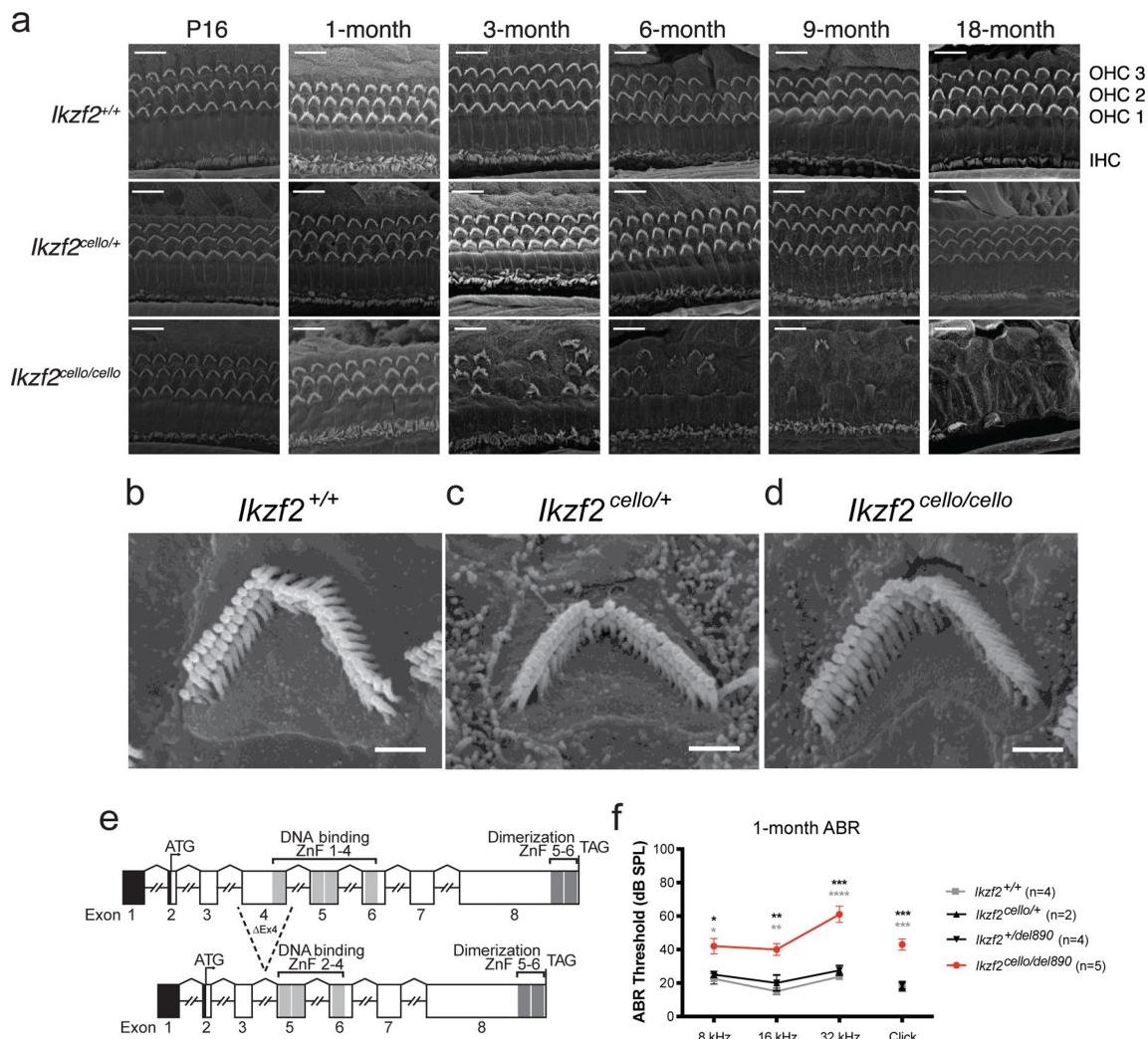
Extended Data Fig. 5 | Scanning electron microscopy of *cello* mice and auditory function of *Ikzf2^cello/del890* compound heterozygotes.
a, Scanning electron micrographs of the organ of Corti of *cello* mice from P16 to 18 months of age. Representative images from the mid-region of the cochlear spiral are shown. Scale bars, 10 μm. n = 3 (P16 *Ikzf2^cello/+*, P16 *Ikzf2^cello/cello*, 1-m *Ikzf2^cello/cello*, 9-m *Ikzf2^+/+*, 18-m *Ikzf2^+/+*, 18-m *Ikzf2^cello/+*, 18-m *Ikzf2^cello/cello*), n = 4 (P16 *Ikzf2^+/+*, 1-m *Ikzf2^+/+*, 3-m *Ikzf2^+/+*, 3-m *Ikzf2^cello/cello*, 6-m *Ikzf2^+/+*, 6-m *Ikzf2^cello/cello*, 9-m *Ikzf2^cello/+*, 9-m *Ikzf2^cello/cello*) and n = 5 (1-m *Ikzf2^cello/+*, 3-m *Ikzf2^cello/+*, 6-m *Ikzf2^cello/+*) biologically independent samples. **b–d**, Scanning electron micrographs of OHC stereocilia bundles of *cello* mice at P16, showing that wild-type *Ikzf2^+/+* (**b**), *Ikzf2^cello+* (**c**) and mutant *Ikzf2^cello/cello* (**d**) mice display overall expected bundle patterning. Images are from the mid-region of the cochlear spiral. Scale bars, 1 μm. n = 3 biologically independent samples for each genotype. **e**, The genomic and domain structure of *Ikzf2^del890*. Black, 5′ untranslated region; light grey, N-terminal DNA-binding domain; dark grey, C-terminal dimerization domain. The *Ikzf2^cello* mutation lies in ZnF6. The *del890* mutation deletes exon 4 and the surrounding intronic sequence. **f**, Averaged ABR thresholds for *Ikzf2^cello/del890* compound heterozygotes at 1 month of age, showing increased thresholds (≥40 dB SPL) at all frequencies tested compared to *Ikzf2^+/+*, *Ikzf2^cello/+* and *Ikzf2^del890/+* control colony mates. Data are mean ± s.e.m. n = 4 (*Ikzf2^+/+*, *Ikzf2^+/del890*), n = 2 (*Ikzf2^cello/+*) and n = 5 (*Ikzf2^cello/del890*) biologically independent samples. *Ikzf2^cello/del890* vs *Ikzf2^+/+*: *P = 0.011 (8 kHz), **P = 0.002 (16 kHz), ****P < 0.0001 (32 kHz), ***P = 0.0001 (click); *Ikzf2^cello/del890* vs *Ikzf2^cello/+*: P = 0.078 (8 kHz), *P = 0.034 (16 kHz), **P = 0.001 (32 kHz), **P = 0.001 (click); *Ikzf2^cello/del890* vs *Ikzf2^+/del890*: *P = 0.025 (8 kHz), **P = 0.009 (16 kHz), ***P = 0.0002 (32 kHz), ***P = 0.0002 (click) (one-way ANOVA with Tukey post hoc test).
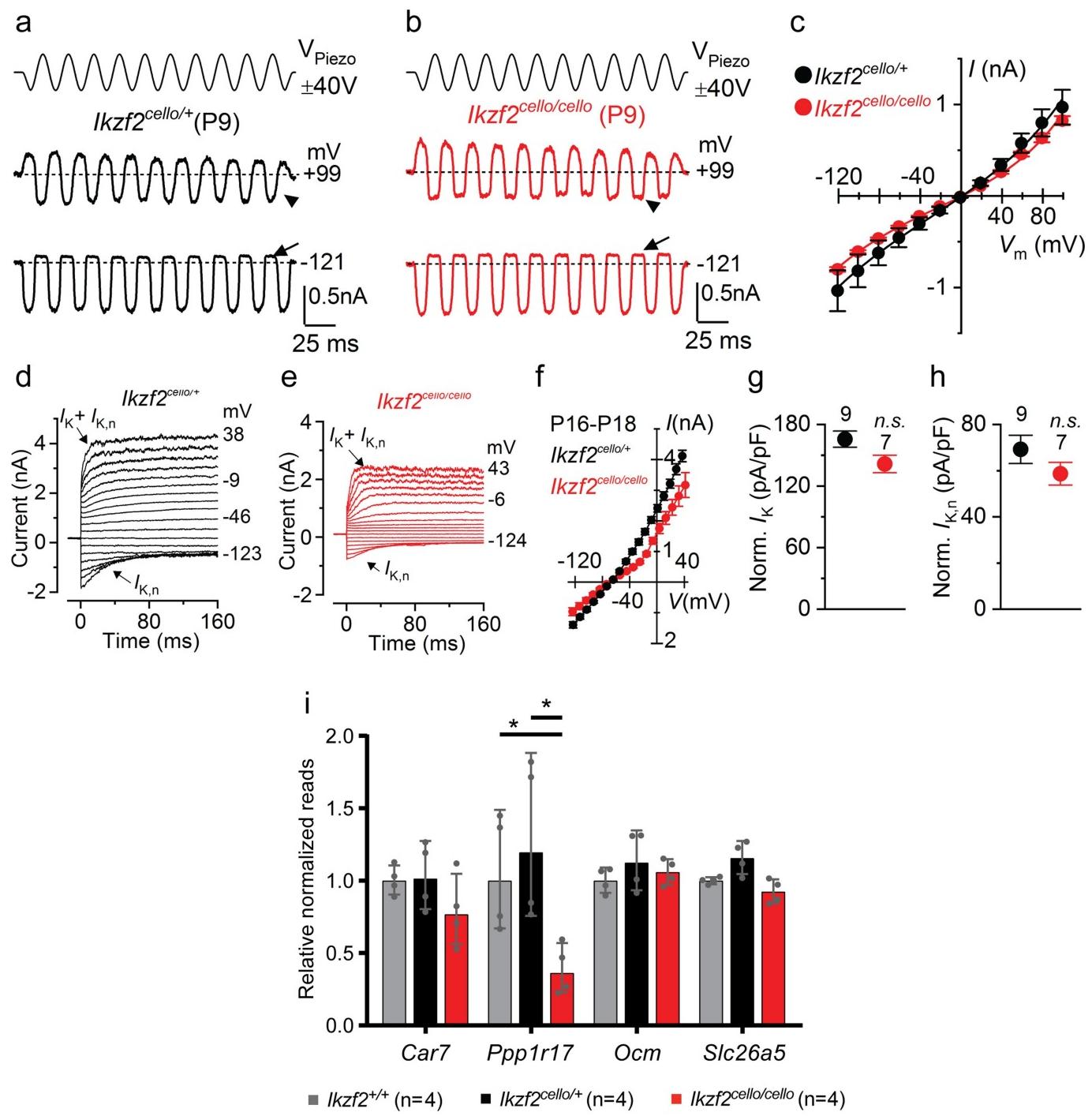
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | The MET and adult-like potassium currents are normal in *Ikzf2^cello* mice. a, b, MET currents were recorded from OHCs of P9 *Ikzf2^{cello/cello}* and *Ikzf2^{cello/+}* (control) littermates. During voltage steps, hair bundles were displaced by applying a 50-Hz sinusoidal force stimuli (the driver voltage to the fluid jet is shown above the traces)[39]. At hyperpolarized membrane potentials (-121 mV), saturating excitatory bundle stimulation (that is, towards the taller stereocilia) elicited a large inward MET current from both *Ikzf2^{cello/+}* and *Ikzf2^{cello/cello}* OHCs, whereas inhibitory bundle stimulation (that is, away from the taller stereocilia) closed the MET channels and reduced the resting current. Because the MET current reverses near 0 mV, it became outward when excitatory bundle stimulation was applied during voltage steps positive to its reversal potential. At positive membrane potentials ($+99$ mV), excitatory bundle stimulation now elicited similar outward MET currents with larger resting amplitudes. Arrows indicate closure of the MET channels (that is, disappearance of the resting current) during inhibitory bundle displacements, arrowheads indicate the larger resting MET current at $+99$ mV compared to -121 mV. **c**, Peak-to-peak current–voltage curves obtained from *Ikzf2^{cello/+}* ($n = 10$ biologically independent samples) and *Ikzf2^{cello/cello}* ($n = 8$ biologically independent samples) OHCs at P9.

The maximal MET current and the resting open probability of the MET channel were found to be similar between the two genotypes. Data are mean $\pm$ s.e.m. **d, e**, Total K^+ currents recorded from P18 *Ikzf2^{cello/+}* control (**d**) and *Ikzf2^{cello/cello}* mutant (**e**) OHCs. The size of the K^+ current, which is mainly due to the negatively activated $I_{K,n}$ (in addition to a small delayed rectifier I_K[15]), was smaller in *Ikzf2^{cello/cello}* OHCs. **f**, Average peak current–voltage relationship for the total K^+ current recorded from the OHCs of *Ikzf2^{cello/+}* ($n = 9$ OHCs from 6 biologically independent animals) and *Ikzf2^{cello/cello}* ($n = 7$ OHCs from 5 biologically independent animals) mice at P16–P18. Data are mean $\pm$ s.e.m. **g, h**, After normalization to the significantly reduced surface area of *Ikzf2^{cello/cello}* OHCs (for this set of experiments: *Ikzf2^{cello/+}*: 14.2 ± 0.4 pF; *Ikzf2^{cello/cello}*: 11.2 ± 0.5 pF; $P < 0.0005$), both the total I_K (**g**) and isolated $I_{K,n}$ (**h**) were not significantly different between the two genotypes at P16–P18 (two-sided Welch's t-test). Data are mean $\pm$ s.e.m. **i**, NanoString validations of genes downregulated in P8 *Ikzf2^{cello/cello}* cochleae at P16, normalized to wild-type reads. Data are mean $\pm$ s.d. ($n = 4$ biologically independent samples per genotype). $*P = 0.038$ (*Ppp17r1* in *Ikzf2^{cello/cello}* vs *Ikzf2^{+/+}*), $*P = 0.037$ (*Ppp17r1* in *Ikzf2^{cello/cello}* vs *Ikzf2^{cello/+}*) (two-sided Welch's t-test).
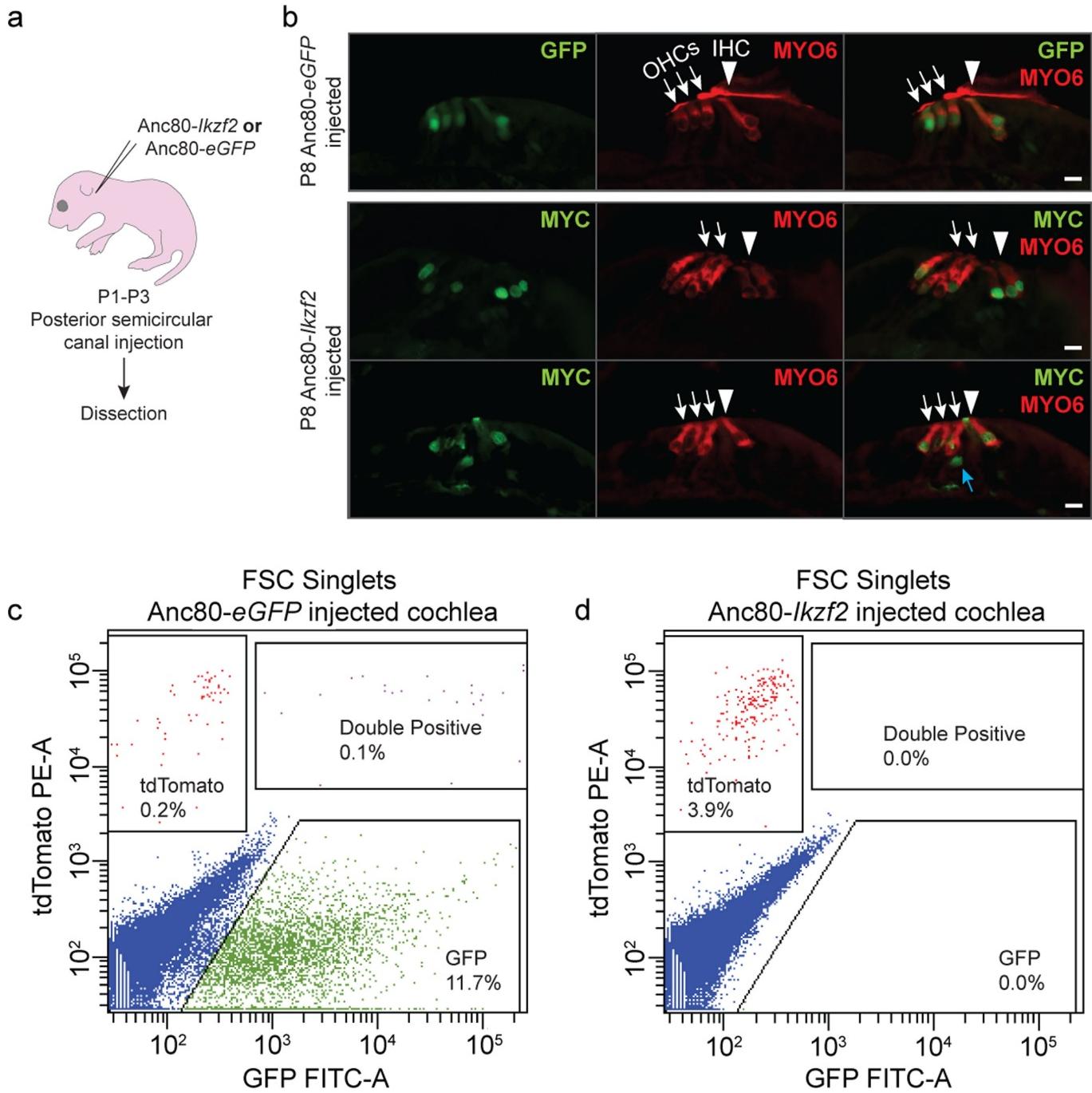
Extended Data Fig. 7 | Transduction of cochlear hair cells using Anc80L65 and hair cell enrichment by flow cytometry. a, Schematic representation of inner ear viral gene delivery via the posterior semicircular canal of CD-1 mice for hair cell marker immunolabelling. **b**, Immunolabelling for GFP in the Anc80-eGFP injected, and MYC in the Anc80-*Ikzf2* injected ears, showing mainly hair cell transduction, although some MYC staining could also be observed in supporting cells (blue arrow). $n = 3$ biologically independent samples per condition. Nuclear MYC staining suggests proper trafficking of the MYC-tagged Helios protein in transduced cells. White arrows indicate OHCs, white arrowheads indicate IHCs. Scale bars, 10 μm. **c, d**, Flow cytometry of dissociated cochlear GFP-positive and tdTomato-positive cells from P8 *Myo15$^{cre/+}$;ROSA26$^{CAG-tdTomato}$* mice injected with either Anc80-eGFP (**c**, 2 mice) or Anc80-*Ikzf2* (**d**, 4 mice). Cells were first gated by forward and side scatter to exclude doublets. For the Anc80-eGFP-transduced cochlear sample, transduced cells were identified based on GFP expression, and hair cells were further identified by tdTomato expression. tdTomato single-positive, GFP single-positive and tdTomato and GFP double-positive cells were collected. For the Anc80-*Ikzf2*-transduced cochlear sample, hair cells were gated based on tdTomato single-positive expression and collected.
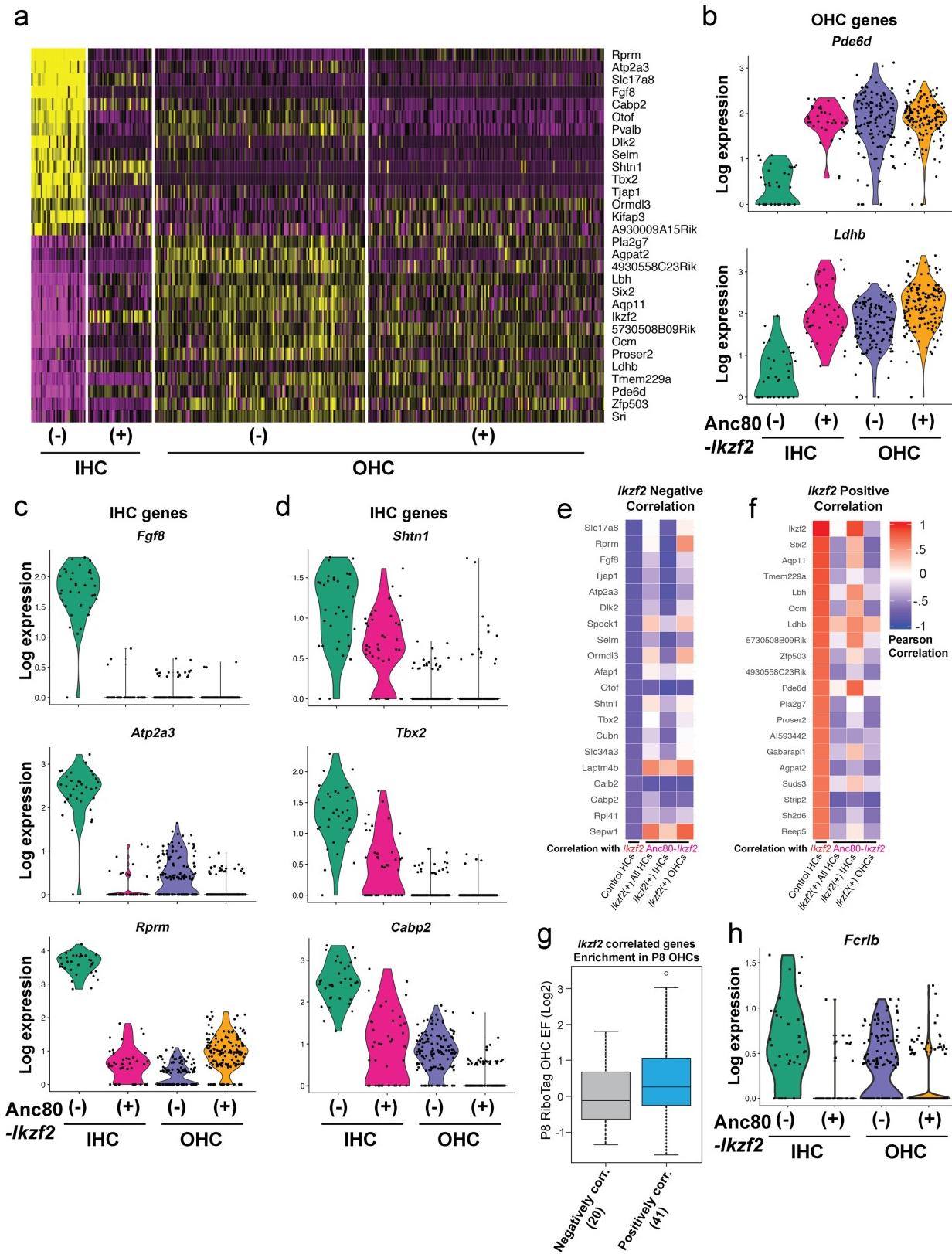
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Transcriptional conversion of Anc80-*Ikzf2*-transduced IHCs. a, Heat map for the top 30 differently expressed genes between all hair cells profiled. Scaled expression values shown as *z*-scores, with yellow indicating higher and purple indicating lower expression than the mean. **b**, OHC enriched genes that are induced in Anc80-*Ikzf2*(+) IHCs. Anc80-*Ikzf2*(−) IHC ($n = 34$) vs Anc80-*Ikzf2*(+) IHC ($n = 40$) FDR: $Pde6d = 2.03 \times 10^{-12}$, $Ldhb = 3.74 \times 10^{-11}$. Dots represent the expression values of individual cells, with width of violins summarizing overall relative distribution of expression. **c**, IHC enriched genes that are highly expressed in control IHCs vs control OHCs, but are significantly reduced in Anc80-*Ikzf2*(+) IHCs. Anc80-*Ikzf2*(−) IHC ($n = 34$) vs Anc80-*Ikzf2*(+) IHC ($n = 40$) FDR: $Fgf8 = 3.30 \times 10^{-14}$, $Atp2a3 = 2.46 \times 10^{-13}$, $Rprm = 2.27 \times 10^{-13}$ (Kruskal–Wallis test followed by post hoc pairwise Wilcoxon ranked sum test adjusted for multiple comparisons). **d**, IHC-enriched genes that show only moderately reduced expression in Anc80-*Ikzf2*(+) IHCs. Anc80-*Ikzf2*(−) IHC ($n = 34$) vs Anc80-*Ikzf2*(+) IHC ($n = 40$) FDR: $Shtn1 = 8.59 \times 10^{-5}$, $Tbx2 = 3.88 \times 10^{-8}$, $Cabp2 = 1.40 \times 10^{-10}$ (Kruskal–Wallis test followed by post hoc pairwise Wilcoxon ranked sum adjusted for multiple comparisons). **e, f**, Top 20 genes negatively (**e**) or positively (**f**) correlated with *Ikzf2* expression in control hair cells, shown alongside corresponding correlations of gene expression within all Anc80-*Ikzf2*-transduced hair cells, Anc80-*Ikzf2*-transduced IHCs, or Anc80-*Ikzf2* transduced-OHCs. See also Extended Data Fig. 9. **g**, Genes that are negatively correlated with *Ikzf2* ($n = 20$, Pearson correlation < −0.6) are not enriched in OHCs at P8 compared to all other genes detected in the RiboTag OHC dataset (background genes, $n = 13,124$). Genes that are positively correlated with *Ikzf2* ($n = 41$, Pearson correlation > 0.6) are significantly enriched in OHCs at P8 compared to background genes ($n = 13,103$) ($P = 0.025$, two-sided Wilcoxon's test). Black line represents median enrichment factor ($\log_2$ fold change), box demarcates first and third quartiles, whiskers demarcate first and third quartile $\pm 1.5 \times$ IQR values, dots represent single outliers. **h**, One of the most differentially expressed genes observed in our scRNA-seq experiment was *Fcrlb*, a gene which encodes an Fc receptor like protein, and the expression of which has not been previously described in the ear. *Fcrlb* is significantly downregulated in Anc80-*Ikzf2*(+) hair cells. Anc80-*Ikzf2*(−) IHC ($n = 34$) vs Anc80-*Ikzf2*(+) IHC ($n = 40$) FDR $= 4.89 \times 10^{-6}$. Anc80-*Ikzf2*(−) OHC ($n = 132$) vs Anc80-*Ikzf2*(+) OHC ($n = 148$) FDR $= 6.88 \times 10^{-8}$ (Kruskal–Wallis test followed by post hoc pairwise Wilcoxon ranked sum test adjusted for multiple comparisons). See also Supplementary Tables 8–11.
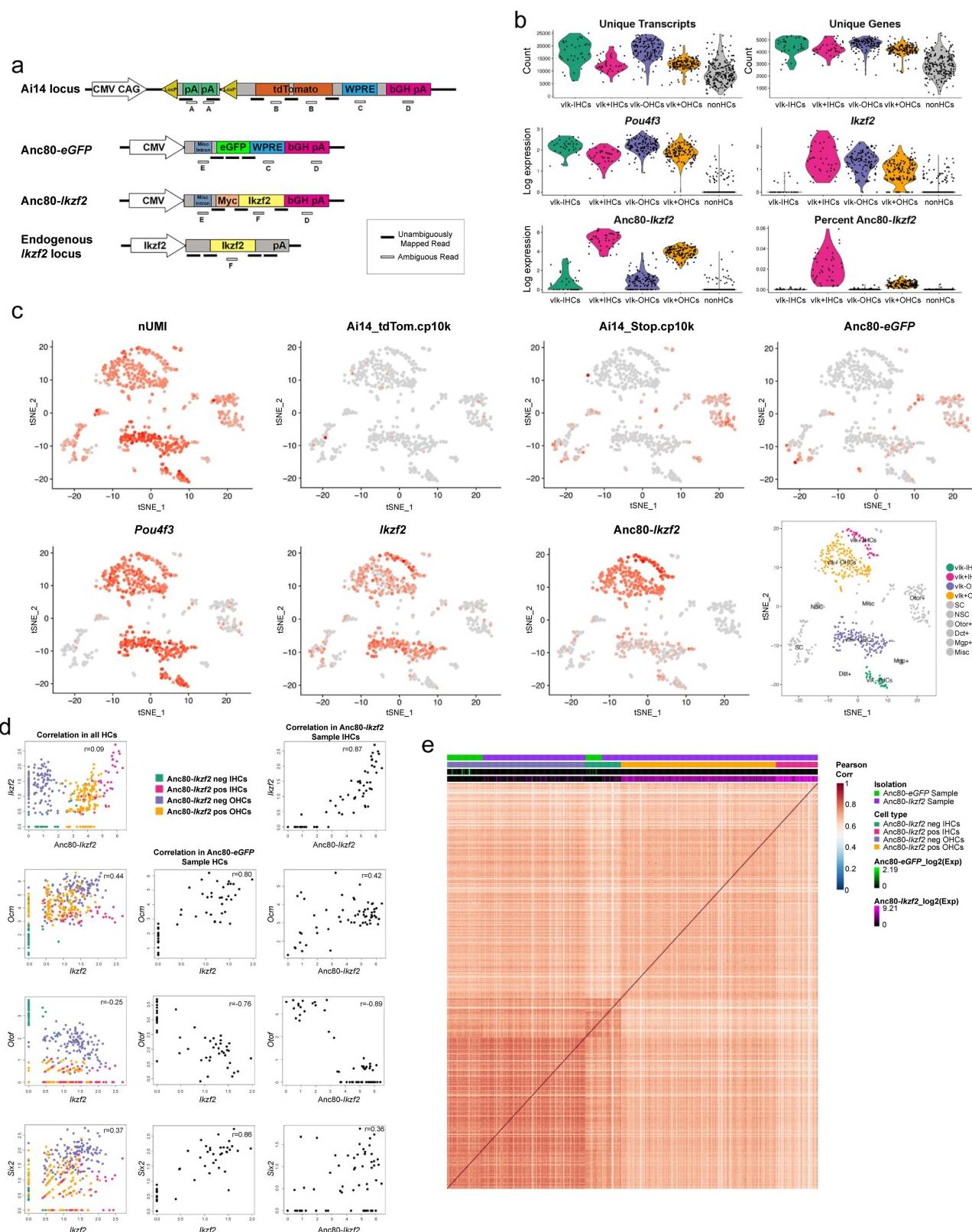
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | scRNA-seq allows for high-resolution discrimination of cell types and their transcriptional changes due to overexpression of *Ikzf2*. a, Custom annotation strategy with theoretical reads mapping to unambiguous regions of the various custom viral loci, as well as those regions that get discarded because of endogenous sequence similarity (that is, ambiguous reads). **b**, Violin plots of the overall scRNA-seq detection metrics, including number of unique molecules detected in each of the major cell type cluster identified (low Anc80-*Ikzf2* expressing IHCs: viral *Ikzf2* (vIk)$^-$ IHCs $n = 34$; low Anc80-*Ikzf2* expressing OHCs: vIk$^-$ OHCs $n = 132$; high Anc80-*Ikzf2* expressing IHCs: vIk$^+$ IHCs $n = 40$; high Anc80-*Ikzf2* expressing OHCs: vIk$^+$ OHCs $n = 140$; and non-HCs: $n = 219$). **c**, FeaturePlots with red showing higher expression across all profiled cells, including cells identified as non-hair cells. Expression from loci captured with custom annotation shown to support cluster identification. A final labelled *t*-SNE plot shows all cells profiled clustered by predicted cell type. Misc, cells from all miscellaneous clusters with fewer than 5 cells; NSC, non-sensory epithelial cell; SC, organ of Corti supporting cell. Other clusters are defined by the highest differentially expressed marker gene. **d**, Pearson correlation scatter plots for selected genes within all profiled hair cells, hair cells from the Anc80-eGFP sample, or IHCs from the Anc80-*Ikzf2* sample. **e**, A Pearson correlation heat map of all hair cells detected showing overall transcriptional similarities between the non-transduced IHCs and OHCs, along with the Anc80-*Ikzf2*-transduced IHCs and OHCs.
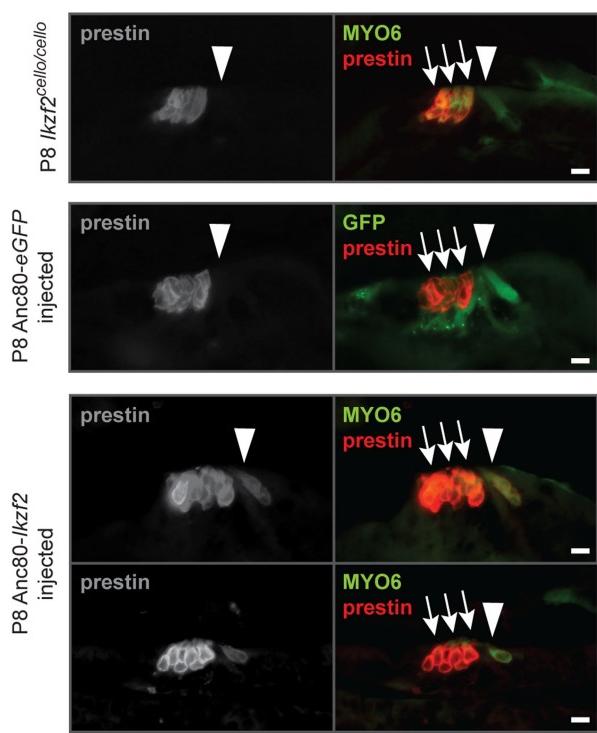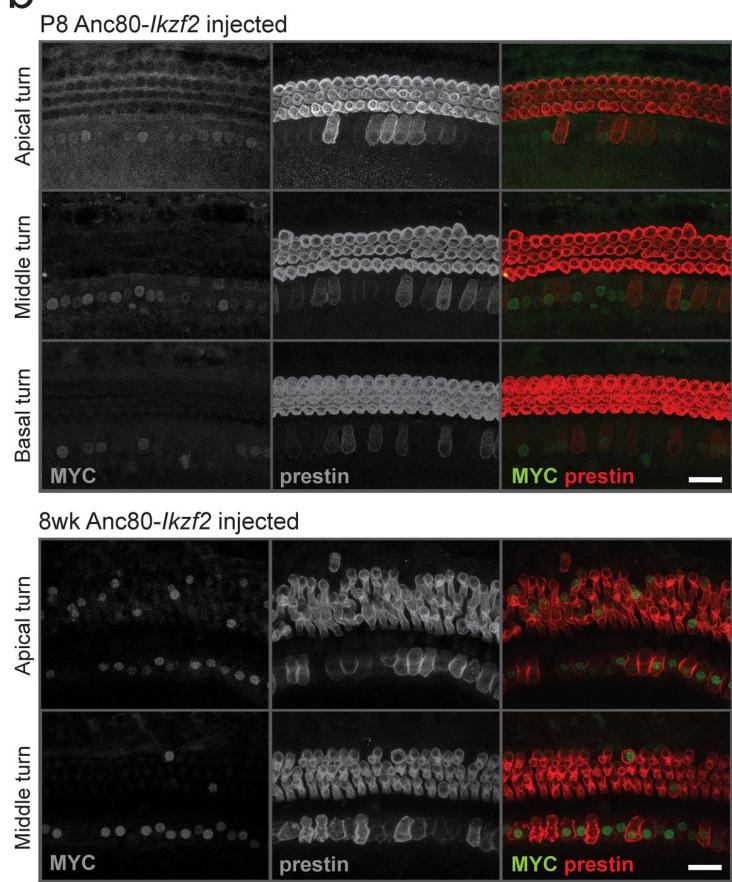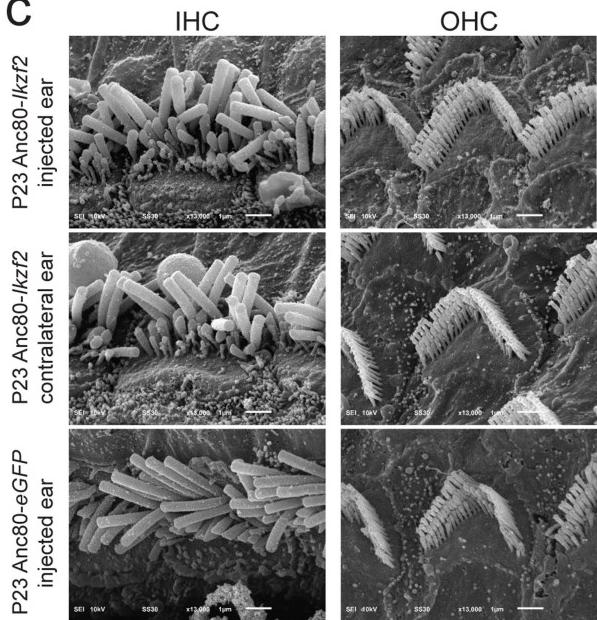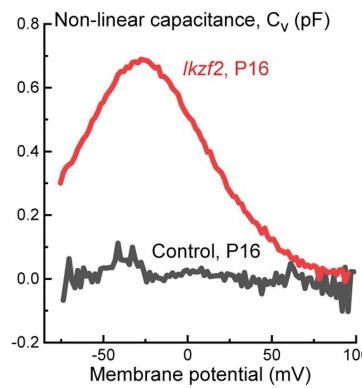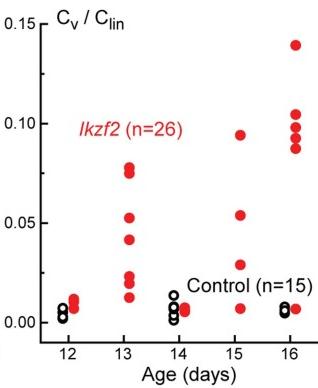
Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | *Ikzf2* overexpression induces prestin expression and electromotility in IHCs but does not affect hair bundle morphology. a, The OHC electromotility protein prestin is expressed in the OHCs of *Ikzf2*$^{cello/cello}$ mutants ($n = 6$ biologically independent samples). In addition, the pattern of prestin expression is not affected by Anc80-eGFP transduction, but is induced in Anc80-*Ikzf2*-transduced IHCs ($n = 3$ biologically independent samples per condition). Scale bars, 10 μm. **b**, Expression of prestin can be seen in Anc80-*Ikzf2*-transduced IHCs as early as P8 and up to 8 weeks of age, and overlaps with MYC staining ($n = 6$ biologically independent samples at P8, $n = 3$ biologically independent samples at 6–8 weeks). Scale bars, 20 μm. **c**, Scanning electron micrographs of IHC and OHC stereocilia bundles of Anc80-*Ikzf2*- and Anc80-eGFP-injected mice at P23 showing expected bundle patterning. Images are from the mid-basal region of the cochlear spiral. Scale bars, 1 μm. Number of biologically independent samples (P16–P23): Anc80-*Ikzf2*-injected cochlea $n = 8$, Anc80-*Ikzf2* contralateral cochlea $n = 6$, Anc80-eGFP-injected cochlea $n = 3$. **d**, Representative traces of the voltage-dependent (nonlinear) component of the membrane capacitance (an electrical signature of electromotility) in the IHCs of Anc80-*Ikzf2*-injected mouse (red) and its non-injected littermate (grey). Mice were injected with Anc80-*Ikzf2* at P2 and recorded at P16. **e**, Normalized maximal nonlinear capacitance in all recorded IHCs of mice injected with Anc80-*Ikzf2* at P2 (red) at different ages after injection and their non-injected littermates (black). Each symbol represents one biologically independent cell, and the total number of cells is indicated in parentheses. Because Anc80-*Ikzf2* transduction is not 100% efficient in the apical turn of the cochlea at the time points tested, some IHCs of Anc80-*Ikzf2*-injected mice do not show prominent nonlinear capacitance, whereas the other IHCs do. In the IHCs with maximal nonlinear capacitance of more than 0.25 pF (due to presumable *Ikzf2* expression), the parameters of the Boltzmann fit were as following (mean ± s.e.m.): $Q_{max} = 0.10 \pm 0.02$ pC; $V_{pk} = -31 \pm 1$ mV; $z = 0.91 \pm 0.02$; $C_{lin} = 11.7 \pm 1.2$ pF; $\Delta C_{sa} = 0.14 \pm 0.07$ pF ($n = 12$). For information on the fitting procedure, see Methods.

LETTER

Linking a cell–division gene and a suicide gene to define and improve cell therapy safety

Qin Liang[1,2,8], Claudio Monetti[1,8], Maria V. Shutova[1], Eric J. Neely[1,2], Sabiha Hacibekiroglu[1,2], Huijuan Yang[1,3], Christopher Kim[1,2], Puzheng Zhang[1], Chengjin Li[1], Kristina Nagy[1,3], Maria Mileikovsky[1], Istvan Gyongy[4], Hoon-Ki Sung[1,5] & Andras Nagy[1,2,6,7]*

Human pluripotent cell lines hold enormous promise for the development of cell-based therapies. Safety, however, is a crucial prerequisite condition for clinical applications. Numerous groups have attempted to eliminate potentially harmful cells through the use of suicide genes[1], but none has quantitatively defined the safety level of transplant therapies. Here, using genome-engineering strategies, we demonstrate the protection of a suicide system from inactivation in dividing cells. We created a transcriptional link between the suicide gene herpes simplex virus thymidine kinase (*HSV-TK*) and a cell-division gene (*CDK1*); this combination is designated the safe-cell system. Furthermore, we used a mathematical model to quantify the safety level of the cell therapy as a function of the number of cells that is needed for the therapy and the type of genome editing that is performed. Even with the highly conservative estimates described here, we anticipate that our solution will rapidly accelerate the entry of cell-based medicine into the clinic.

Most randomly integrated transgenes show variegated expression[2–4]. To achieve reliable expression, we generated a transcriptional link between a cell-division locus (CDL), which is a gene that is essential for a cell to divide or to survive, and a drug-inducible suicide system (SU), which resulted in a CDL–SU allele (Fig. 1a). Therefore, the expression of the CDL and the drug-inducible suicide system are tightly linked and, if required, either only the dividing cells or all of the cells can be arrested or eliminated by treatment with the drug that induces the suicide system.

We consider a cell population a safe-cell batch when all cells within that batch contain a functional suicide system. The safe-cell level (SCL) is the number of therapeutic batches in which there is expected to be only one non-safe batch (Fig. 1b), such that one in a thousand gives a SCL of 1,000 and one in a million has a SCL of 1,000,000. We used various in vitro and in vivo experiments as well as mathematical modelling to define the SCL as the function of the number of cells needed for therapy.

From a list of CDL candidates (Supplementary Table 1), we chose *CDK1* and *HSV-TK.007*[5] (*TK*) as prototypes. The absence of CDK1 causes a block in the G2 to M transition of the cell cycle, and other CDKs are not able to rescue this deficiency[6–8]. TK has been extensively used for cell ablation[9] and its mechanism of action in the presence of its clinically approved prodrug ganciclovir (GCV), is well-characterized[10]. *CDK1* is not expressed in non-dividing cells[8]. Therefore, TK is not expressed from the *CDK1–TK* allele, eliminating the potential effect of its immunogenicity[11].

To generate a transcriptional link between *CDK1* and *TK* (Fig. 1c), we inserted *TK* into the 3′ untranslated region of *Cdk1* in mouse C2 embryonic stem (ES) cells[12] (Extended Data Fig. 1a–c) and of *CDK1* in human H1[13] (Extended Data Fig. 2a–f) and human CA1[14] (Extended Data Fig. 3a–e) ES cells. We determined the optimal GCV concentration for the heterozygous *Cdk1–TK*-expressing mouse and heterozygous *CDK1–TK*-expressing human ES cells in vitro (Extended Data

Figs. 1d, 2h) as well as for controlling the growth of the teratomas that were generated using these lines. Human ES cells implanted in immunodeficient NOD/SCID/IL2Rγ (NSG) mice and mouse ES cells implanted in isogenic C57BL/6N recipient mice resulted in teratoma formation with the expected efficiency (Extended Data Figs. 1e, 2g). At a volume of 500 mm³ (day 0), we administered GCV daily by intraperitoneal injection for up to four weeks. GCV rendered the C2 ES-cell-derived teratomas dormant, without growth rebound following the treatment (Fig. 1d and Extended Data Fig. 4a, b). H1-derived teratomas responded similarly although occasionally, repeated GCV treatments were required to stabilize the teratoma size (Fig. 1f and Extended Data Fig. 4d). The decrease in teratoma size after GCV readministration indicates regained proliferation of quiescent or slow-dividing cells following drug withdrawal. Consequently, TK is expressed in dividing cells and induces subsequent GCV sensitivity. The volume of the human teratomas frequently increased in the later phases; however, in agreement with previous reports[15,16], this was the result of cyst formation (Extended Data Fig. 4e) and not solid tissue growth.

The induced long-term dormancy of teratomas was encouraging, but was also unexpected given that such a large tissue (approximately 10⁹ cells[17]) could contain numerous cells that might be capable of escaping the suicide system through different types of mutations. Within the well-encapsulated teratoma, however, these presumably resistant cells (escapees) could have been eliminated by the bystander-killing effect of the TK–GCV system[18].

To further investigate the capacity of the safe-cell system to control cell proliferation, we performed a breast cancer transplantation assay using safe-cell mammary epithelial tumour cells[19]. Upon isogenic transplantation, we observed that after a delayed period (approximately 100 days), heterozygous safe-cell tumours became resistant to GCV and they continued to grow in the presence of the drug (Extended Data Fig. 5a, b).

To identify escapees that appeared during the expansion of heterozygous safe-cell ES cells, we designed an in vitro experiment that mitigated the bystander-killing effect (Extended Data Fig. 6a, b) and characterized the mechanisms by which resistance occurred in eight independent clonal escapee lines obtained from 120 million cells. To determine whether the mechanism by which resistance occurred was caused by large genomic changes or *Cdk1–TK* locus-specific mutations, we analysed the copy number of *Cdk1*, the TK transgene and six endogenous genes that are found on chromosome 10 (Extended Data Fig. 6d, f). Only one escapee (E3 in Extended Data Fig. 6f) contained the *TK* gene (Extended Data Fig. 6c). We did not detect mutations in the coding region of either *Cdk1* or *TK* (data not shown); however, the expression level of this allele was reduced and rendered cells GCV-resistant (Extended Data Fig. 6e). Another escapee, E5, was the result of a regional deletion that included the *Cdk1–TK* locus, which led to a more than 18.5-Mb hemizygous region in the wild-type chromosome (Extended Data Fig. 6d, f). Diploidy (copy number of 2) in the ten genes

[1]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada. [2]Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada. [3]Department of Physiology, University of Toronto, Toronto, Ontario, Canada. [4]School of Mathematics and Maxwell Institute, The University of Edinburgh, Edinburgh, UK. [5]The Hospital for Sick Children Research Institute, Toronto, Ontario, Canada. [6]Australian Regenerative Medicine Institute, Monash University, Melbourne, Victoria, Australia. [7]Department of Obstetrics & Gynaecology, University of Toronto, Toronto, Ontario, Canada. [8]These authors contributed equally: Qin Liang, Claudio Monetti. *e-mail: nagy@lunenfeld.ca
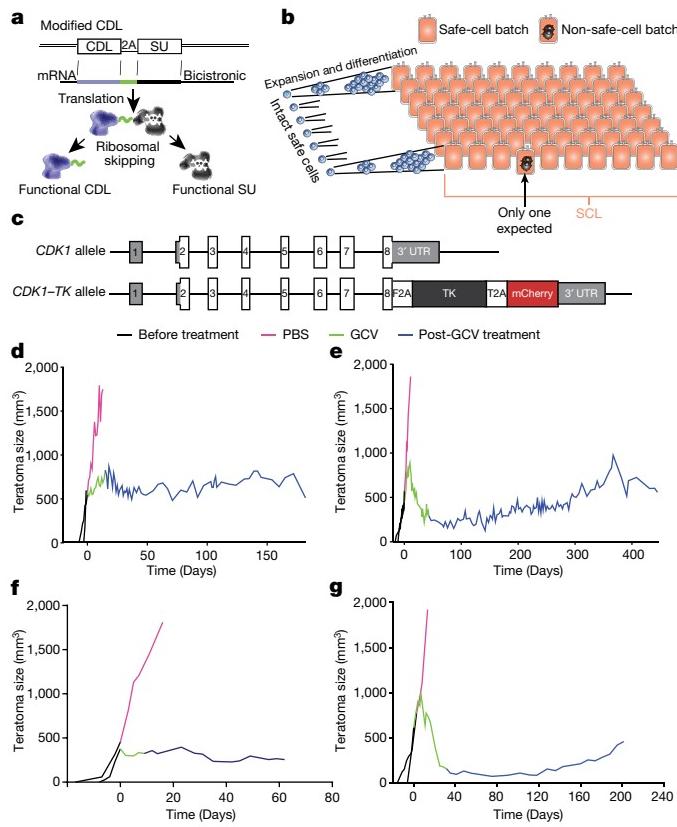
Fig. 1 | The concept, the definition, realization and properties of the safe-cell system. a, The suicide gene is placed into a cell-division essential locus (CDL), resulting in a bicistronic mRNA that is translated into two proteins; a cell division essential factor and a drug-inducible suicide factor. **b**, Visual representation of the SCL defined by one non-safe-cell batch out of many batches. **c**, The link between the prototype drug-inducible suicide system (*HSV-TK*) and the prototype CDL (*CDK1*). 3′ UTR, 3′ untranslated region. **d**, Representative growth of teratomas formed by mouse *Cdk1–TK/Cdk1* cells, when the recipient mice were treated with PBS or GCV. **e**, Representative growth of teratomas formed by human *CDK1–TK/CDK1* ES cells, when the recipient mice were treated with PBS or GCV. **f**, Representative growth of teratomas in mice formed by mouse *Cdk1–TK/Cdk1–TK* ES cells. **g**, Representative growth of teratomas in mice formed by human *CDK1–TK/CDK1–TK* ES cells. **d–g**, Experiments were repeated multiple times (see Extended Data Fig. 4) with similar results.

Fig. 2 | Modelling of the safe-cell system and calculating the SCL. a, Genotype transition matrix considered in the modelling and Monte Carlo simulations. **b**, Visual illustration of safe-cell and non-safe-cell batch formation during cell expansion. QC, quality control. **c**, The function between therapeutic cell number and SCL was determined by Monte Carlo simulation (data points) for different initial cell genotypes. cn, cell number. Solid lines show the approximated linear regression on the 'close to linear' parts of the functions. Monte Carlo simulations were run *n* times for each genotype ($n = 80 \times 10^6$ for CDL–SU/wild-type; $n = 10^8$ for CDL–SU homozygous; $n = 10^7$ for CDL–SU/null heterozygous; $n = 13.5 \times 10^6$–90.5×10^6 for (two-CDL)-SU homozygous). The bars on certain data points represent the 95% confidence intervals of the SCL estimates.

and the lack of TK in the remaining six resistant clones suggested that these were formed by diploid loss of heterozygosity (dLOH), which was probably caused by mitotic recombination or chromosomal non-disjunction, leading to homozygosity of the wild-type *Cdk1* allele in a diploid form. These data indicate that dLOH is the dominant mechanism by which the *Cdk1–TK* allele is lost in heterozygous ES cells, consistent with a study of mouse *Aprt*[+/−] heterozygous cells, in which dLOH accounted for 78% of the loss of gene function events[20].

To mitigate the generation of escapees by dLOH, we established both mouse and human ES cell lines that were homozygous for the *Cdk1–TK* and *CDK1–TK* alleles, respectively (Extended Data Figs. 1a–c, 2a–h, 3a–e). As expected, we were unable to identify any escapees (Extended Data Fig. 6b). Homozygous ES-cell-derived teratomas behaved similarly to teratomas derived from heterozygous ES cells; a brief GCV treatment was sufficient to render the teratomas dormant (Fig. 1f, g and Extended Data Fig. 4c, f). As in human heterozygous ES-cell-derived lines, cyst formation also occurred in human homozygous ES-cell-derived teratomas (Extended Data Fig. 4f).

To test the limits of our system, we generated homozygous safe-cell mouse mammary tumour cells and transplanted the cells into wild-type isogeneic C57BL/6N recipient mice. We found that the size of tumours was reduced following transplantation and their growth was

restrained after GCV administration (Extended Data Fig. 5c). The observed growth rebound following GCV withdrawal was not surprising as slow-dividing and quiescent tumour-prone cells survive GCV administration and can start proliferating in the absence of the drug. Nevertheless, even in this non-clinical, extreme situation in which a tumour cell line is used for cell transplantation, the homozygous safe-cell system is capable of controlling tumour growth.

Because we were unable to identify any escapees that appeared in homozygous safe-cell ES cells, we used Monte Carlo simulation to estimate the odds of escapees in this scenario. The model considers three types of mutations that could potentially influence the function of the CDL–SU link (Extended Data Fig. 7a). Type-1 mutations (su1) render the suicide system non-functional while keeping the linked CDL operational. Type-2 mutations (su2) eliminate both the suicide system and CDL functionality through epigenetic or genetic changes to the entire locus, including a hemizygous LOH-dependent mechanism. Type-3 mutations (su3) remove a functional CDL–SU allele by dLOH.

To estimate the probabilities of type-1, -2 and -3 mutations (P_1, P_2 and P_3, respectively) per cell generation (Extended Data Fig. 7a), we designed our own experiment (Supplementary Table 4) and also used published data[21–24]. We used the values $P_1 = P_2 = 10^{-6}$ and $P_3 = 2 \times 10^{-5}$ per cell per division, all of which are intentional overestimates. Consequently, our calculated SCLs represent underestimates, being equal to or lower than the actual SCL.

In silico, we subsequently generated a sufficient number of cell batches that were expanded from a single cell with an intact suicide
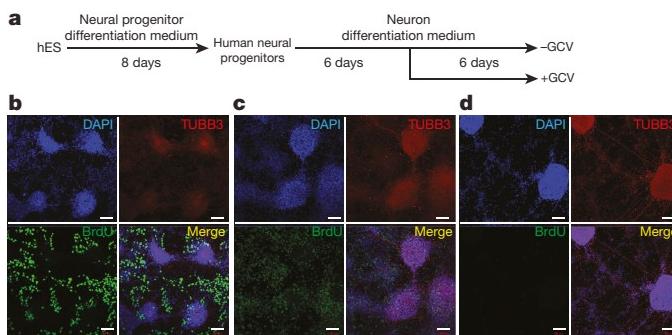
Fig. 3 | The safe-cell system eliminates only proliferating cells.
a, Schematic. **b–d**, Human *CDK–TK/CDK1–TK* ES cells are differentiated into neural epithelial progenitors (**b**) and subsequently into neurons without (**c**) or with (**d**) GCV treatment. **b**, Neural epithelial progenitors consist of a mixed culture of proliferating neural epithelial progenitors and non-proliferating differentiated neurons that were treated with GCV. Scale bars, 100 μm. This experiment was repeated three times with similar results.

system. During each doubling, the model permits allele transitions (Extended Data Fig. 7b) that determine the transition graph (Fig. 2a), reflecting the genotype change that could occur during cell expansion. For the homozygous *CDK1–TK/CDK1–TK* simulation, we initiated the batch production from an SU/SU cell with two intact suicide-system alleles, whereas for the heterozygous *CDK1–TK/CDK1^{WT}* simulation, the initiating cell was SU/su1, because the su1 allele is functionally equivalent to the *CDK1^{WT}* allele. For the compound heterozygous *CDK1–TK/CDK1^{null}* simulation, the initiating cell was SU/su2, as the su2 allele is the same as a *CDK1^{null}* allele (Fig. 2a).

A batch of cells is considered to be a safe-cell batch if it does not contain any escapees (Fig. 2b). On the basis of the frequency of getting a non-safe-cell batch of cells determined by the Monte Carlo simulation, we calculated the SCL as the function of the number of cells that is needed for a therapeutic cell batch. Figure 2c shows these functions for the different initiating cell genotypes described above.

The number of cells that is required for cell therapy is disease-specific and is estimated to range between 10^5 (for example, the eye[25,26]) and 10^{10} (for example, the heart[27]) cells. The genotype scenarios presented in Fig. 2c show that a single TK insertion (*CDK1–TK/CDK1^{WT}* or *CDK1–TK/CDK1^{null}*) gives a low SCL. By contrast, a homozygous TK insertion into *CDK1* significantly increases the SCL and brings the safety level into a clinically relevant range. However, for diseases that

require a larger number of therapeutic cells (10^8 to 10^{10} cells), the SCL provided even by the homozygous CDL–SU is insufficient (SCL < 10). For this disease category, we propose the use of a homozygous modification of two different CDLs. In this scenario, our Monte Carlo simulation showed a strong increase in SCL (SCL > 10^6 for all clinically relevant batch volumes, Fig. 2c).

We observed that the logarithm of SCL, as a function of the logarithm of the cell number, is very close to a linear function when the SCL is above 10 and the cell number is in the clinically relevant range. Therefore, for an estimation of SCL, we applied linear regression to these segments (Fig. 2c). Using these approximates, the calculation of SCLs becomes very simple, while retaining the desired underestimates (Fig. 2c): for one CDL, this equates to SCL = 10^9/cn, and for two CDLs this equates to SCL = 10^{16}/cn, where 'cn' is the number of cells needed for a therapeutic cell batch.

During the production of therapeutic cells, some are lost during differentiation or expansion. Therefore, the efficiency of cell production should be accurately estimated and the cell numbers that are needed to generate a therapeutic batch should be corrected accordingly.

In future cell therapies, if allogeneic cells are desired or HLA haplo-banks[28] of pluripotent cell lines are available, generation of off-the-shelf cell batches would be advantageous. This would require the production of large pools of cells that, following quality control, would be aliquoted into therapeutic batches. We calculated the effect of aliquoting on SCL using both mathematical and Monte Carlo modelling approaches (equation (1), Supplementary Information and Extended Data Fig. 8a). Notably, aliquoting resulted in an approximately fivefold decrease in SCL using both approaches.

Quality control should be performed on every cell pool to ensure that the originating cell was a safe-cell and consequently, the SCL calculation is correct. To this end, we grew several batches from a single, homozygous safe-cell ES cell. At the early phase of expansion, we verified that both *CDK1–TK* alleles were expressed and intact using flow cytometry, allele-specific PCR (Extended Data Fig. 8b–e) and sequencing of the TK coding region.

Both mouse and human ES cells with homozygous modifications of *CDK1* have a normal ES cell morphology, self-renewing capacity and ability to differentiate (Extended Data Fig. 9). Additionally, using in vitro neural differentiation, we demonstrate the selective killing of dividing cells by the safe-cell system. Following a brief GCV treatment, all mitotically active cells were eliminated whereas non-dividing cells were spared (Fig. 3). This ability could represent a valuable safety measure before transplantation of cells into a patient.
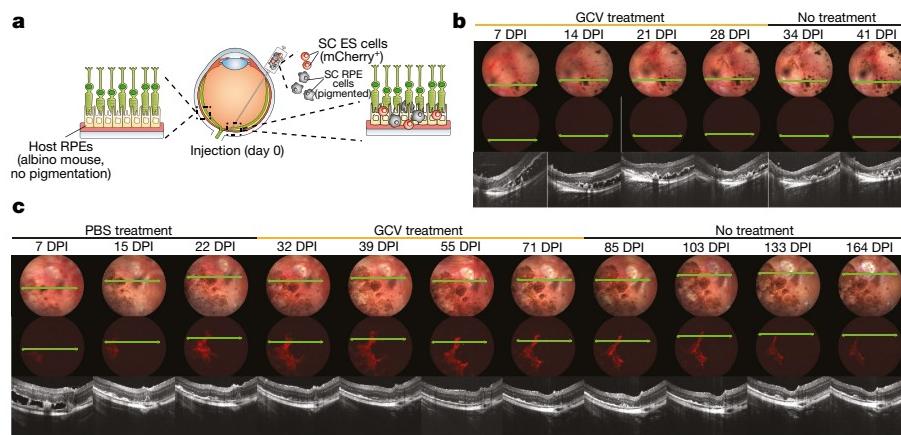
Fig. 4 | An in vivo proof of principle study shows the safe-cell system in action. a, A 3:1 mixture of human homozygous safe-cell ES-cell-derived RPE cells and homozygous safe-cell ES cells were subretinally injected into NSG mice and imaged using fundoscopy and optical coherence tomography throughout GCV or PBS treatment. **b**, Fundoscopy, optical coherence tomography and fluorescence imaging of the eyes of mice that received GCV treatment (4 weeks). **c**, Fundoscopy, optical coherence tomography and fluorescence imaging of the eyes of a mouse that received PBS treatment (3 weeks) and developed an actively growing ES-cell-derived lesion (mCherry+ cells). d.p.i., days post-injection.
b, **c**, Experiments were repeated multiple times with similar results (Extended Data Fig. 10).

To simulate a clinical cell transplantation scenario gone awry, we injected a 3:1 mixture of human homozygous safe-cell retinal pigment epithelium (RPE) cells and human homozygous safe-cell ES cells (that were tagged with mCherry) into the subretinal space in the eyes of NSG mice. Among four injected eyes, we did not observe any ES cell derivatives when GCV was administered 24 h post-injection and for 28 days, as a preventative measure (Fig. 4a, b and Extended Data Fig. 10a, b). However, cell growth was detectable in six eyes that received PBS as an initial treatment (Fig. 4c and Extended Data Fig. 10d–f). Notably, even when GCV administration was delayed three weeks post-injection and cell growth was present, the homozygous safe-cell system efficiently arrested the ES-cell-derived component of the graft; only non-dividing cells remained (Fig. 4c and Extended Data Fig. 10e, f). This experiment illustrates the ability of the safe-cell system to selectively eliminate proliferating cells after cell transplantation. Neither the initial nor the delayed GCV treatment affected the RPE graft or the integrity of the surrounding retinal tissue (Fig. 4b, c and Extended Data Fig. 10).

No therapy is without risk. Our safe-cell system (concept and genome-editing approach) provides a definition of risk and a quantification of the safety level as a function of the number of cells that is needed for any given cell therapy. We contend that the risks associated with the safe-cell system are sufficiently low to provide an indispensable component of prospective cell therapies. Our approach to assessing and quantifying the safety of cell-based therapies will be critical for informed decision-making by regulators, clinicians and patients while advancing modern medicine.

Online content

1. Li, W. & Xiang, A. P. Safeguarding clinical translation of pluripotent stem cells with suicide genes. *Organogenesis* **9**, 34–39 (2013).
2. Dobie, K., Mehtali, M., McClenaghan, M. & Lathe, R. Variegated gene expression in mice. *Trends Genet.* **13**, 127–130 (1997).
3. Yagyu, S., Hoyos, V., Del Bufalo, F. & Brenner, M. K. Multiple mechanisms determine the sensitivity of human-induced pluripotent stem cells to the inducible caspase-9 safety switch. *Mol. Ther. Methods Clin. Dev.* **3**, 16003–16008 (2016).
4. Kotini, A. G., de Stanchina, E., Themeli, M., Sadelain, M. & Papapetrou, E. P. Escape mutations, ganciclovir resistance, and teratoma formation in human iPSCs expressing an HSVtk suicide gene. *Mol. Ther. Nucleic Acids* **5**, e284 (2016).
5. Preuß, E. et al. TK.007: A novel, codon-optimized HSVtk(A168H) mutant for suicide gene therapy. *Hum. Gene Ther.* **21**, 929–941 (2010).
6. Diril, M. K. et al. Cyclin-dependent kinase 1 (Cdk1) is essential for cell division and suppression of DNA re-replication but not for liver regeneration. *Proc. Natl Acad. Sci. USA* **109**, 3826–3831 (2012).
7. Santamaría, D. et al. Cdk1 is sufficient to drive the mammalian cell cycle. *Nature* **448**, 811–815 (2007).
8. Satyanarayana, A. et al. Genetic substitution of Cdk1 by Cdk2 leads to embryonic lethality and loss of meiotic function of Cdk2. *Development* **135**, 3389–3400 (2008).
9. Fillat, C., Carrió, M., Cascante, A. & Sangro, B. Suicide gene therapy mediated by the herpes simplex virus thymidine kinase gene/ganciclovir system: fifteen years of application. *Curr. Gene Ther.* **3**, 13–26 (2003).
10. Tomicic, M. T., Thust, R. & Kaina, B. Ganciclovir-induced apoptosis in HSV-1 thymidine kinase expressing cells: critical role of DNA breaks, Bcl-2 decline and caspase-9 activation. *Oncogene* **21**, 2141–2153 (2002).
11. Berger, C., Flowers, M. E., Warren, E. H. & Riddell, S. R. Analysis of transgene-specific immune responses that limit the in vivo persistence of adoptively transferred HSV-TK-modified donor T cells after allogeneic hematopoietic cell transplantation. *Blood* **107**, 2294–2302 (2006).
12. Gertsenstein, M. et al. Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos. *PLoS ONE* **5**, e11260 (2010).
13. Thomson, J. A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
14. The International Stem Cell Initiative. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat. Biotechnol.* **25**, 803–816 (2007).
15. Damjanov, I. & Andrews, P. W. Teratomas produced from human pluripotent stem cells xenografted into immunodeficient mice - a histopathology atlas. *Int. J. Dev. Biol.* **60**, 337–419 (2016).
16. Wu, S. M. & Hochedlinger, K. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. *Nat. Cell Biol.* **13**, 497–505 (2011).
17. Bianconi, E. et al. An estimation of the number of cells in the human body. *Ann. Hum. Biol.* **40**, 463–471 (2013).
18. Mesnil, M. & Yamasaki, H. Bystander effect in herpes simplex virus-thymidine kinase/ganciclovir cancer gene therapy: role of gap-junctional intercellular communication. *Cancer Res.* **60**, 3989–3999 (2000).
19. Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol. Cell. Biol.* **12**, 954–961 (1992).
20. Cervantes, R. B., Stringer, J. R., Shao, C., Tischfield, J. A. & Stambrook, P. J. Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc. Natl Acad. Sci. USA* **99**, 3586–3590 (2002).
21. Larson, J. S., Yin, M., Fischer, J. M., Stringer, S. L. & Stringer, J. R. Expression and loss of alleles in cultured mouse embryonic fibroblasts and stem cells carrying allelic fluorescent protein genes. *BMC Mol. Biol.* **7**, 36 (2006).
22. Mortensen, R. M., Conner, D. A., Chao, S., Geisterfer-Lowrance, A. A. & Seidman, J. G. Production of homozygous mutant ES cells with a single targeting construct. *Mol. Cell. Biol.* **12**, 2391–2395 (1992).
23. Koike, H. et al. Efficient biallelic mutagenesis with Cre/*loxP*-mediated inter-chromosomal recombination. *EMBO Rep.* **3**, 433–437 (2002).
24. Yusa, K. et al. Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. *Nature* **429**, 896–899 (2004).
25. Schwartz, S. D. et al. Human embryonic stem cell-derived retinal pigment epithelium in patients with age-related macular degeneration and Stargardt's macular dystrophy: follow-up of two open-label phase 1/2 studies. *Lancet* **385**, 509–516 (2015).
26. Mandai, M. et al. Autologous induced stem-cell-derived retinal cells for macular degeneration. *N. Engl. J. Med.* **376**, 1038–1046 (2017).
27. Chong, J. J. H. et al. Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. *Nature* **510**, 273–277 (2014).
28. Barry, J., Hyllner, J., Stacey, G., Taylor, C. J. & Turner, M. Setting up a haplobank: issues and solutions. *Curr. Stem Cell Rep.* **1**, 110–117 (2015).

Author contributions Q.L. designed and conducted most of the experiments, including targeting of mouse and human ES cells, teratoma analysis, mammary gland tumour analysis, neural in vitro differentiation, mutation rate calculation, as well as analysing the data and writing the manuscript. C.M. designed and conducted experiments, designed and constructed the *CDK1–TK* targeting cassette, targeted mouse and human ES cells, analysed safe-cell escapees, performed in vitro differentiation assays, analysed data, composed figures and wrote the manuscript. M.V.S. conducted the Monte Carlo simulations and analysed the data. I.G. performed the mathematical modelling. E.J.N. targeted human ES cells, differentiated RPEs and wrote the manuscript. S.H. performed the eye experiment and analysed the data. H.Y. performed in vitro differentiation into endoderm and analysed the data. C.K. performed in vitro differentiation into mesenchymal stem cells and analysed the data. P.Z. performed Southern blots. C.L. performed animal teratoma experiments. K.N. performed animal teratoma experiments, analysed the data and edited the manuscript. M.M. targeted human ES cells. H.-K.S. analysed teratoma histology. A.N. conceived and supervised the study, designed experiments, analysed the data and wrote the manuscript.

METHODS

Generation of targeting vectors. Targeting vectors were generated by DNA synthesis, molecular cloning, recombineering and the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs).

Generation of CRISPR–Cas9 vectors. pX330-U6-Chimeric_BB-CBh-hSpCas9 was a gift from F. Zhang (Addgene plasmid 42230)[29]. Guide sequences for CRISPR–Cas9 were analysed using the online CRISPR design tool (http://crispr.mit.edu). Guide sequence for mouse *Cdk1* targeting: TAAGAAGATGTAGCCCTC. Guide sequence for human *CDK1* targeting: CTATCTGTTGACATAACATA.

Mouse ES cell culture. C57BL/6N C2 ES cells were grown at 37 °C in 95% air, 5% CO$_2$ on mouse embryonic fibroblasts (MEFs) obtained from TgN(DR4)1Jae/J mice (The Jackson Laboratory, 003208) at all times except for one passage on gelatinized tissue-culture plates before aggregation[30]. Two types of medium were used. The first medium, FBS–DMEM ES cell medium, was used for gene targeting. This medium consisted of high-glucose DMEM supplemented with 15% FBS (previously shown to support germline chimaera generation), 2 mM GlutaMAX, 1 mM Na pyruvate, 0.1 mM non-essential amino acids (NEAA), 50 U ml^{-1} penicillin–streptomycin (all Thermo Fisher Scientific), 0.1 mM 2-mercaptoethanol (Sigma-Aldrich) and 1,000 U ml^{-1} LIF prepared with LIF-producing plasmid[31]. 2). The second medium, KSR+2i medium was used for 2–4 passages before the generation of ES cell chimaeras[12]. KSR+2i medium consisted of high-glucose DMEM medium supplemented with 15% knockout serum replacement (KSR) (Thermo Fisher Scientific), 1 mM Na pyruvate, 0.1 mM NEAA, 0.1 mM 2-mercaptoethanol, 2 mM GlutaMAX, 50 U ml^{-1} penicillin–streptomycin, 500 U ml^{-1} LIF, 5 mg ml^{-1} insulin (Thermo Fisher Scientific), 1 μM of the mitogen-activated protein kinase inhibitor PD0325901 (StemGent) and 3 μM of the glycogen synthase kinase-3 inhibitor CHIR99021 (StemGent). Cells were fed daily and passaged when they reached a confluency of 70–80%. Then, 0.05% trypsin-EDTA (Thermo Fisher Scientific) was used for the passaging of cells grown in FBS–DMEM and accutase (STEMCELL Technologies) was used for cells grown in KSR+2i medium. Cells were tested (negative) for mycoplasma contamination but were not authenticated.

Mouse ES cell targeting. In brief, 50,000 mouse C57BL/6N C2 ES cells were transfected with 2 μg DNA (mouse target vector I or II, 1.5 μg; CRISPR vector, 0.5 μg) using JetPrime for transfection (Polyplus). The cells were selected for G418 resistance (160 μg ml^{-1}) starting 48 h after transfection. Resistant clones were picked independently and replicated in 96-well plates for freezing and genotyping using PCR. PCR-positive clones were expanded, frozen in multiple vials and genotyped by Southern blotting.

Selection cassette excision in mouse ES cells. Correctly targeted ES cell clones were transfected with episomal-hyPBase (for mouse target vector I) or pCAGGs-NLS-Cre-IRES-puromycin (for mouse target vector II). Then, 2–3 days after transfection, cells were trypsinized and plated clonally (1,000–2,000 cells per 10-cm plate). mCherry$^+$ clones were picked and transferred to 96-well plates into independent wells and genotyped by PCR and Southern blotting to confirm the excision event. The junctions of the removal region were PCR-amplified, sequenced and confirmed to be intact and without any frameshift mutations. GCV (Sigma-Aldrich) to test for TK activity was used at a final concentration of 1 μM.

Human ES cell culture. Human CA1 and H1 ES cells were cultured on Geltrex (Thermo Fisher Scientific) using mTeSR1 medium (STEMCELL Technologies) containing 50 U ml^{-1} penicillin–streptomycin (Thermo Fisher Scientific). Cells were passaged using TrypLE Express (Thermo Fisher Scientific) and were subsequently plated in mTeSR medium containing 10 μM ROCK inhibitor (Selleckchem) for 24 h. Cells were tested (negative) for mycoplasma contamination but were not authenticated.

Human ES cell targeting. For human targeting vectors I and II, six million human ES cells were electroporated using a Neon Transfection System (Thermo Fisher Scientific) with protocol 14 (pulse voltage, 1,200 mV; pulse width, 20 ms; pulse number, 2) with 24 μg DNA (target vector, 18 μg; CRISPR vector, 6 μg). After transfection, cells were plated on four 10-cm plates. G418 selection at 30 μg ml^{-1} or puromycin selection at 0.75 μg ml^{-1} was initiated 48 h after transfection. Independent colonies were picked to 96-well plates and each plate was duplicated for further growth and genotyping with PCR. PCR-positive clones were expanded, frozen in multiple vials and genotyped with Southern blotting. For human target vector III targeting, 10 million human ES cells were electroporated using Neon protocol 14 with 40 μg DNA (human target vector III, 30 μg; CRISPR vector, 10 μg) and plated in four 10-cm plates. After 3–4 days of transfection, cells that were double-positive for mCherry and eGFP were sorted into one well of a 96-well plate. After recovery from fluorescence-activated cell sorting (FACS), cells were dissociated and plated clonally (1,000–2,000 cells per 10-cm plate). Next, clones were picked independently, replicated and transferred to 96-well plates for freezing and genotyping with PCR. PCR-positive clones were expanded, frozen in multiple vials and genotyped by Southern blotting.

Selection cassette excision in human ES cells. In brief, one million correctly targeted ES cell clones were electroporated with 2 μg episomal-hyPBase-IRES-puro

(for human target vector I) or 2 μg episomal-Cre-IRES-puro (for human target vector II) using Neon protocol 14. Once the cells were confluent in six-well plates, mCherry$^+$ cells were sorted into one well of a 96-well plate by FACS. After recovery, cells were dissociated and plated clonally (1,000–2,000 cells per 10-cm plate). Clones were picked and transferred to 96-well plates into independent wells and genotyped by PCR and Southern blotting to confirm the excision event. The junctions of the removal region were PCR-amplified, sequenced and confirmed to be intact and without frameshift mutations.

PCR genotyping. For all PCR reactions, 2× Taq PCR master mix (Biomart) was used. Genomic DNA from human cell pellets was extracted using the DNeasy Blood & Tissue Kit (Qiagen). The primer pairs and conditions used for each reaction are listed in Supplementary Table 2.

Southern blotting. In brief, 10 μg of genomic DNA was extracted from PCR-positive clones, digested with ScaI-HF overnight, resolved by 0.6–0.7% gel electrophoresis, and transferred to Hybond N+ (GE Healthcare). The following probes were labelled with ^{32}P and used to hybridize with the membrane (around 25 ng probe per ml hybridization solution). Human *CDK1* genomic probe: PCR-amplified with primers h*CDK1*-Probe6-F and h*CDK1*-Probe6-R. Mouse *Cdk1* genomic probe: PCR-amplified with primers 647302FWD and 647302REV. mCherry probe: the entire length of mCherry. eGFP probe: the entire length of eGFP. TK–mCherry probe: cut from h*CDK1*-PB-neo-TK-mCherry with Bsu36I and SgrAI: 1,092 bp, gel-purified.

Mice. The CD-1 (ICR) (Charles River) outbred albino mouse stock was used as embryo donors for aggregation with ES cells and as pseudopregnant recipients. Six-to-ten-week-old C57BL/6NCrl mice (Charles River) were used as the host for teratoma assays with mouse C2 ES cells. Six-to-ten-week-old C57BL/6NCrl or B6N-Tyr_c N4/Crl#493 (Charles River) mice were used as the host for mammary fat pad transplantation of mammary epithelial cells. Six-to-ten-week-old NSG/J#5557 mice (Jackson Laboratories) were used as the host for teratoma assays with human H1 or CA1 ES cells. FVB/N-Tg(MMTV-PyVT)634Mul/J mice were a gift from members of W. Muller's laboratory[19], and the backcross to B6J background was done by members of A. Pawson's laboratory. Animals were maintained on a 12-h light/dark cycle and provided with food and water ad libitum in individually ventilated units (Techniplast) in the specific-pathogen free facility at The Centre for Phenogenomics (TCP). All procedures involving animals were performed in compliance with the Animals for Research Act of Ontario and the Guidelines of the Canadian Council on Animal Care. Animal protocols performed in this study were approved by the Toronto Centre for Phenogenomics Animal Care Committee, chaired by A. Jurisicova and L. Phaneuf. The number of animals used in different experiments was determined in accordance with similar studies in the field; owing to the nature of most experiments, blinding was impossible, because the results are visible at the time of analysis. Animals were allocated randomly when possible.

Generation of chimaeras and mouse lines. Morula aggregations were performed as previously described[30]. Chimaeras were identified at birth by the presence of black eyes and later by coat pigmentation. Male chimaeras with more than 50% ES cell contribution to coat colour were bred with CD-1 females to identify germline transmitters. The transmitter was then bred with C57BL/6NCrl females and pups were confirmed by genotyping to obtain *Cdk1–TK/Cdk1* mice. *Cdk1–TK/Cdk1* MMTV-PyMT males were generated by breeding MMTV-PyMT (B6) males and *Cdk–TK/Cdk1* females. *Cdk1–TK/Cdk1* MMTV-PyMT and *Cdk–TK/Cdk1–TK* MMTV-PyMT female mice were generated by breeding *Cdk1–TK/Cdk1* MMTV-PyMT males and *Cdk1–TK/Cdk1* females.

Teratoma assay. Matrigel Matrix High Concentration (Corning) was diluted 1:3 with cold DMEM medium on ice. Then, 1–5 million mouse ES cells or 5–10 million human ES cells were suspended in 100 μl of Matrigel–DMEM and injected subcutaneously into one or both dorsal flanks of C57BL/6NCrl mice (for mouse C2 ES cells) and NSG/J#5557 mice (for human H1 and CA1 ES cells). Teratomas formed 2–4 weeks after injection. Teratoma size was measured with callipers and volume was calculated using the formula $V = (L \times W \times H)\pi/6$. GCV or PBS treatment was performed through daily intraperitoneal injections (50 mg kg^{-1}) with varying treatment durations. At the end of treatment, mice were euthanized and tumours were dissected and fixed in 4% paraformaldehyde for histological analysis.

Breast cancer transplantation assay. *Cdk1–TK/Cdk1* MMTV-PyMT and *Cdk1–TK/Cdk1–TK* MMTV-PyMT female mice developed mammary gland tumours between three and six months of age. Mammary epithelial tumorigenic cells were isolated from developed tumours by digestion in 10× collagenase–hyaluronidase (STEMCELL Technologies), and dilution to 1× with medium consisting of DMEM/F12 (Thermo Fisher Scientific), 10% FBS and 50 U ml^{-1} penicillin–streptomycin for 1 h in 37 °C. The digested cells were washed and pelleted with DMEM/F12 and 10% FBS four times, and plated in CnT-PRIME epithelium culture medium (CELLnTEC Advanced Cell Systems) on plates coated with 0.1% gelatin (Sigma-Aldrich). Without passaging, primary mammary epithelial cells were dissociated and resuspended in PBS at 10,000 cells per μl, and 50 μl (500,000 cells)

was transplanted into each mouse by intraductal injection after making a small abdominal skin incision as previously described[32]. Tumour measurement and PBS or GCV treatment were the same as described in the teratoma assay.

Differentiation of human ES cells into RPE cells. RPE differentiation was performed as previously described[33] with minor changes. Human ES cells were plated on Geltrex-coated six-well plates and cultured in feeder-free conditions with mTeSR medium until confluency was reached and the colonies lost their tight borders (7–10 days). Next, the medium was replaced with differentiation medium (basal media with 13% KSR) and changed every 2–3 days. The basal medium consisted of KO-DMEM supplemented with 50 U ml^{-1} penicillin–streptomycin, 1 mM Na pyruvate, 0.1 mM NEAA, 2 mM GlutaMAX and 0.1 mM 2-mercaptoethanol. Initial pigmentation was observed approximately three weeks after the switch to differentiation medium. Clusters of RPE cells were manually picked and transferred to a Geltrex-coated 24-well plate (three clusters per well) when they were large enough (around 1 mm in diameter) for enrichment and the medium was changed to RPE medium, which consisted of basal medium with 5% FBS, 7% KSR and 10 ng ml^{-1} bFGF (Peprotech).

Differentiation of human ES cells into definitive endoderm. Definitive endoderm differentiation was performed using the STEMdiff Trilineage Differentiation Kit (StemCell Technologies) and characterized by immunostaining for SOX17 and FOXA2 (Supplementary Table 3).

Differentiation of human ES cells into pharyngeal pouch endoderm. Differentiation into pharyngeal pouch endoderm was performed as previously described[34] with the only modification being that the induction from ES cells to definitive endoderm is one day shorter than reported.

Differentiation of human ES cells into mesenchymal stem cells and subsequent adipogenic, osteogenic and chondrogenic differentiation. ES cells were cultured in mTeSR medium for two days. Next, cells were induced into early mesoderm progenitor cells with STEMdiff Mesenchymal Induction Medium (STEMCELL Technologies) for four days and then maintained in MesenCult-ACF Medium (STEMCELL Technologies). Cells were continually passaged into six-well plates precoated with MesenCult-ACF Attachment Substrate (STEMCELL Technologies) to derive early mesenchymal progenitor cells. At day 21, the mesenchymal stem cells (MSCs) showed a fibroblast-like morphology and the culture medium was changed every three days. For adipogenic differentiation, MSCs at a density of 20,000 cells per well were plated with MesenCult-ACF Attachment Substrate and cultured with MesenCult-ACF Medium for two days. Adipogenesis was induced using the StemPro Adipogenesis Differentiation Kit (Thermo Fisher Scientific). After 21 days, lipid droplets were visualized using Oil Red O (Sigma). For osteogenic differentiation, ES-derived MSCs at a density of 50,000 cells per well were plated with MesenCult-ACF Attachment Substrate and cultured with MesenCult-ACF Medium for two days. Osteogenesis was induced using the StemPro Osteogenesis Differentiation Kit (Thermo Fisher Scientific). After 21 days, calcium deposition was visualized using Alizarin Red (Sigma-Aldrich). For induction of chondrogenic differentiation, ES-derived MSCs were centrifuged in 15-ml conical tubes at 500g for 5 min to create cell pellets with 5,000,000 cells per pellet. Chondrogenesis was induced using the StemPro Chondrogenesis Differentiation Kit (Thermo Fisher Scientific). After 21 days, cartilage was visualized using Alcian Blue (Sigma-Aldrich). Differentiation medium was changed every three days.

Differentiation of human ES cells into beating cardiomyocytes. Cardiomyocyte differentiation was performed using the STEMdiff Cardiomyocyte Differentiation Kit (STEMCELL Technologies).

Differentiation of human ES cells into neuronal progenitors and neurons. To differentiate human ES cells into neuronal progenitors, human ES cells were plated at 50–100,000 cells per cm^2 in 1:1 DMEM/F12: Neurobasal (Thermo Fisher Scientific), 0.5× N2 supplement (made in-house, 1.92 mg ml^{-1} putrescine, 2.376 μg ml^{-1} progesterone, 3.6 μM selenium, 10 mg ml^{-1} apo-transferrin, 0.75% BSA, 20 g ml^{-1} insulin), 0.5× B27 supplement with vitamin A (Thermo Fisher Scientific), 2 mM GlutaMAX, 0.1 mM β-mercaptoethanol, 50 U ml^{-1} penicillin–streptomycin, 10 μM SB431542 (Selleckchem), 100 nM LDN193189 (Selleckchem) (10 μM ROCK inhibitor was added overnight only and was subsequently removed). Cells were maintained in this condition for eight days and the medium was changed every other day. Next, neuronal progenitors were dissociated with accutase and plated at a density of 5 × 10^4 cells per cm^2 on laminin (Sigma-Aldrich, 1 μl for 1cm^2, diluted in 250 μl PBS without Ca and Mg) in fast neuron differentiation medium, 1:1 DMEM-F12:Neurobasal, 1× B27 supplement with vitamin A, 5 μM DAPT (Selleckchem), 2 mM GlutaMAX, 0.1 mM β-mercaptoethanol, 50 U ml^{-1} penicillin–streptomycin. Medium was changed every three days. Then, 10 μM GCV was added six days after neuron differentiation and kept for six days. After five days of GCV treatment, 10 μM BrdU (Sigma-Aldrich) was added and the cultures were fixed after six days of GCV treatment and were immunostained for BrdU and β-tubulin III (Supplementary Table 3).

Flow cytometry analysis and FACS. Flow cytometry and FACS experiments were both performed and analysed by the Lunenfeld-Tanenbaum Research Institute

flow cytometry facility. FACS was performed using the ASTRIOS EQ cell sorter. Flow cytometry was performed using the GALLIOS flow cytometer and evaluated using Kaluza Analysis Software (Beckman Coulter). Samples were gated for live single cells using forward scatter, side scatter and DAPI staining. Wild-type and single-colour samples of the same cell type as the experimental samples were used as negative controls and compensation calculations. Human ES cell samples were single-cell sorted using StemFlex Medium (Thermo Fisher Scientific) and 10 μM ROCK inhibitor.

Immunostaining. Cells fixed in 4% PFA were blocked and permeabilized with 5% goat serum, 1 M glycine and 1% Triton X-100 (all Sigma-Aldrich) in PBS without Ca and Mg or animal-free blocker (Vector Laboratories) with 1% Triton X-100 in milliQ water. All of the primary antibody information can be found in Supplementary Table 3. Staining was visualized using a Zeiss LSM780 confocal microscope.

Histology analysis. Paraffin embedding, paraffin block sectioning, and haematoxylin and eosin staining were performed by the Pathology Core of The Centre for Phenogenomics.

qPCR. Gene expression analyses were completed as follows: RNA extraction by GenElute Mammalian Total RNA Miniprep Kit (Sigma), reverse transcription with QuantiTect Reverse Transcription Kit (Qiagen), qPCR with SensiFAST SYBR No-Rox Kit (Bioline) on Bio-Rad CFX Real-Time Systems (Bio-Rad) and analysis with Bio-Rad CFX Manager 3.1. All information regarding primers and probes for the TaqMan qPCR analyses can be found in Supplementary Table 2. For copy number analysis, the reactions were performed using TaqMan Genotyping Master Mix (Thermo Fisher Scientific) and CFX Real-Time Systems, and were analysed by CopyCaller Software v.2.1 (Thermo Fisher Scientific).

Luria and Delbruck assay. The Luria and Delbruck assay was performed as previously described[35]. CDK1–TK homozygous 3C cells were single-cell plated in a 96-well plate using FACS. Subsequently, 21 single-cell-derived cultures were grown to an average of five million cells per culture, and the numbers of single-positive cells in each culture were analysed using flow cytometry. The mutation rate was calculated using the previously described equation[36] available at https://www.wolframalpha.com/.

In vivo transplantation of RPEs. CDK1–TK homozygous 3C ES cells were transfected with the PB-CAGGs-mCherry-pA plasmid and sorted for highly expressing cells. Then, 40,000 3C-derived RPEs only or 30,000 3C-RPEs and 10,000 mCherry-tagged 3C ES cells were injected subretinally with 0.5%/0.5% (wt/vol) hydrogel blend of hyaluronan and methylcellulose (HAMC) in HBSS. PBS or GCV (50 mg kg^{-1}) treatments were started the day after cell injection or as stated in the Figures, and were given every other day through intraperitoneal injections. Monitoring by fundoscopy and optical coherence tomography was performed on the day after transplantation and then once a week.

SCL calculations. To established the rate of mutation in the CDK1–TK allele, we used our targeted human H1 CDK1–TK–mCherry/CDK1–TK–eGFP dichromatic cell line, as most mutations in either CDK1–TK allele result in monochromatic cells. We grew 21 parallel cultures from a single dichromatic cell to an average of 5 × 10^6 cells per culture (>22 consecutive doublings) and determined the number of monochromatic cells in the culture using flow cytometry. Next, we applied Luria–Delbruck fluctuation analysis[35,36] to calculate the sum of the $P_1 + P_2 + P_3$ probabilities in the two CDK1–TK alleles. We found that the mutation rate of losing mCherry was 9.05 × 10^{-6} per cell per division while the mutation rate of losing eGFP was similar at 7.68 × 10^{-6} per cell per division (Extended Data Fig. 8a, b). To further validate the probabilities of these various mutations, we also analysed published studies that focused on these events. In mouse ES cells, the mutation rate ($P_1 + P_2 + P_3$) of changing from a dichromatic to a monochromatic phenotype in the Rosa26 locus (mouse chromosome 6) was 1 × 10^{-5} per cell, per division[21]. Similarly, another study calculated the mutation rate of gene function loss in the Gdf9 locus (mouse chromosome 11) to be 2.3 × 10^{-5} events per cell per division[37]. Furthermore, the probability of the type 3 mutation, P_3 alone, has been calculated as 1 × 10^{-5}, 7.2 × 10^{-6} and 8.5 × 10^{-6} in three different studies[22–24] by performing high-G418 selection in mouse ES cells. The $P_1 + P_2$ mutation rate has also been estimated in the human HPRT locus on the X chromosome to be 1.7–6 × 10^{-7} by Luria–Delbruck fluctuation analysis[38], and 5 × 10^{-6} through mutation frequency analysis in population datasets[39]. Next, we performed Monte Carlo simulations to establish the SCL of cell batches derived from different SU genotypes. On the basis of both published data and our own, we used the values $P_1 = P_2 = 10^{-6}$ and $P_3 = 2 × 10^{-5}$ per cell per division; all of which are intentional overestimates. An ES cell population was considered to be a mix of mutant and non-mutant cells with reference to the CDL-SU locus (or loci). All possible mutations were categorized into three different types: type 1, when only the SU part of the locus becomes non-functional (su1 allele); type 2, when both the CDL and SU become non-functional (su2 allele); type 3, when any of the above occurs as a result of LOH (Fig. 2a, b). Back mutations, such as su1 to SU, su2 to SU or su2 to su1, were not considered, because of their extremely low probabilities (Fig. 2c). Back mutations,

such as SU/su1 to SU/SU, SU/su2 to SU/SU and su1/su2 to su1/su1, were considered as a part of the more frequent LOH process. P_1, P_2 and P_3 were designated the probabilities of each mutation type, respectively. We distinguished between two types of P_3: P_{3mr} (probability LOH occurred through mitotic recombination), where both daughter cells survive; and P_{3cnd} (probability LOH occurred through chromosomal non-disjunction), where one of the daughter cells with the single remaining copy of the chromosome is likely to die[40]. On the basis of these probabilities, matrices of transitions between all possible genotypes within one or two CDL systems were constructed (Fig. 2c, https://github.com/mashutova/failsafe). With each division cycle (d), all cells within the population except cells with the su2/su2 genotype, were allowed to divide. Genotypes with su1/su2 and su1/su1 were considered escapees and the simulation initiated from one non-mutant cell. $n(g_1,d)$ was the number of cells of genotype g_1 at doubling d, and $P(g_1,g_2)$ was the probability of transition from genotype g_1 to g_2. In each doubling, the number of cells changing genotype from g_1 to g_2 was determined through random sampling from a binomial distribution with parameters $2n(g_1,d-1)$ and $P(g_1,g_2)$. We used a Poisson approximation of a binomial distribution to work with ultra-low $P(g_1,g_2)$ values. For each division, the number of cells of each genotype was assessed, and the simulation proceeded until the first escapee was detected. For each starting genotype, we performed more than 10 million simulations and obtained a distribution of the number of doublings (d) from the detection of the first escapee. On the basis of these data, we generated a function of SCL (overall number of trials divided by number of trials with escapees) over cell population size (2^d) (Fig. 2d). Because all graphs contain almost linear regions, we used linear models to extrapolate them to high SCL values. To obtain linear regression lines, we used only simulated points from the linear-like part of the graph ($R^2 > 0.999$) with 95% confidence intervals less than 1,000. To obtain a conservative boundary for the SCL, we used only the lowest confidence interval values to build linear regressions. To analyse the outcome from the aliquoting of the pool of safe-cell cells possibly containing escapees, using probability modelling we developed the following formula to calculate the drop of SCL (for details see equation (1))

$$\frac{SCL_{aP}}{SCL_a} \approx \frac{1}{A\left(1 - \sum_{k=0}^{m-1} 2^{-k-1}\left(\frac{A-1}{A}\right)^{2^k}\right)} \tag{1}$$

To reduce the complexity of the model, we considered only one escapee event in the pool, as the possibility of two independent escapees occurring in a pool is low in the quasi-linear phase of SCL. Nevertheless, we tested the effect of this omission on the drop of SCL due to aliquoting using Monte Carlo simulations. We performed 10 million independent trials for a doubling of 20 and a doubling of 27, and obtained a distribution of the number of escapees for each of them. Through

randomly sampling a number of escapees from each trial to the A aliquots, we calculated the number of 'bad' aliquots containing one or more escapees (A_b). To calculate a new SCL of the population after aliquoting, where SCL_p is the SCL of the original population, we used the formula $A \times SCL_P/\text{mean}(A_b)$. The drop in SCL was measured in silico and was compared with the value that we obtained from the equation obtained from the probability model.

Code availability. The code used in this manuscript is publicly available at https://github.com/mashutova/failsafe.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

29. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
30. Behringer, R., Gertsenstein, M., Nagy, K. V. & Nagy, A. *Manipulating the Mouse Embryo: A Laboratory Manual* 4th edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2013).
31. Mereau, A., Grey, L., Piquet-Pellorce, C. & Heath, J. K. Characterization of a binding protein for leukemia inhibitory factor localized in extracellular matrix. *J. Cell Biol.* **122**, 713–719 (1993).
32. Park, J., Kusminski, C. M., Chua, S. C. & Scherer, P. E. Leptin receptor signaling supports cancer cell metabolism through suppression of mitochondrial respiration in vivo. *Am. J. Pathol.* **177**, 3133–3144 (2010).
33. Klimanskaya, I. et al. Derivation and comparative assessment of retinal pigment epithelium from human embryonic stem cells using transcriptomics. *Cloning Stem Cells* **6**, 217–245 (2004).
34. Sun, X. et al. Directed differentiation of human embryonic stem cells into thymic epithelial progenitor-like cells reconstitutes the thymic microenvironment in vivo. *Cell Stem Cell* **13**, 230–236 (2013).
35. Luria, S. E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511 (1943).
36. Capizzi, R. L. & Jameson, J. W. A table for the estimation of the spontaneous mutation rate of cells in culture. *Mutat. Res.* **17**, 147–148 (1973).
37. Luo, G. et al. Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nat. Genet.* **26**, 424–429 (2000).
38. Monnat, R. J. Jr. Molecular analysis of spontaneous hypoxanthine phosphoribosyltransferase mutations in thioguanine-resistant HL-60 human leukemia cells. *Cancer Res.* **49**, 81–87 (1989).
39. Green, M. H. L., O'Neill, J. P. & Cole, J. Suggestions concerning the relationship between mutant frequency and mutation rate at the *hprt* locus in human peripheral T-lymphocytes. *Mutat. Res.* **334**, 323–339 (1995).
40. Biancotti, J.-C. et al. The in vitro survival of human monosomies and trisomies as embryonic stem cells. *Stem Cell Res.* **9**, 218–224 (2012).
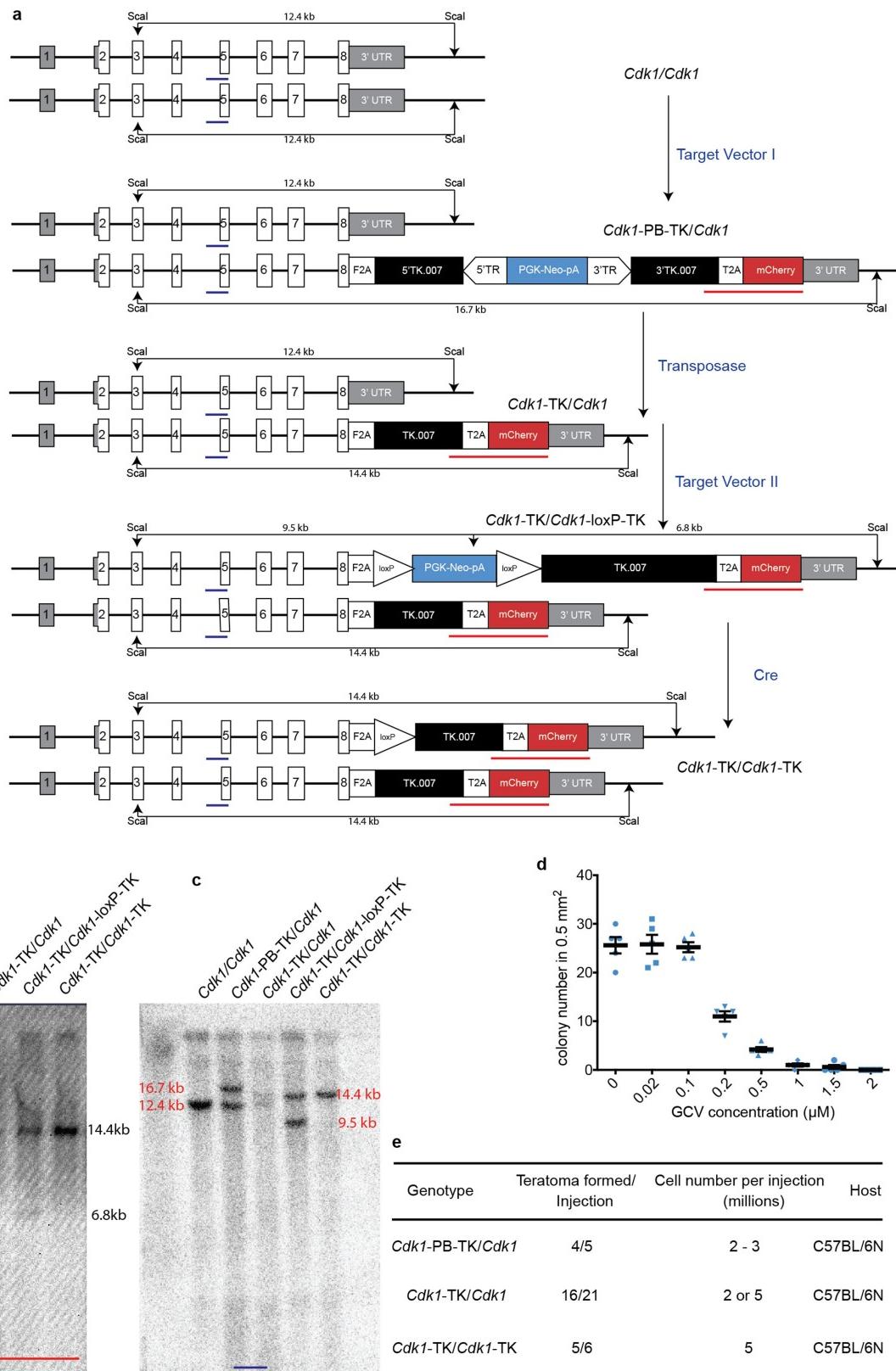
Extended Data Fig. 1 | Generation, genotyping and characterization of mouse C57BL/6N C2 *Cdk1–TK/Cdk1* and *Cdk1–TK/Cdk1–TK* ES cells. a, Summary of the targeting steps used to generate mouse C2 *Cdk1–TK/Cdk1* and *Cdk1–TK/Cdk1–TK* ES cells. **b**, Southern blot genotyping with internal TK–mCherry probe. **c**, Southern blot genotyping with mouse *Cdk1* genomic probe. **d**, In vitro GCV dose–response killing curve of mouse C2 *CDK1–TK/CDK1* ES cells. Data are mean ± s.d., $n = 3$. **e**, Teratoma formation efficiency of mouse C2 *Cdk1–TK-PB/Cdk1*, *Cdk1–TK/Cdk1* and *Cdk1–TK/Cdk1–TK* ES cells.
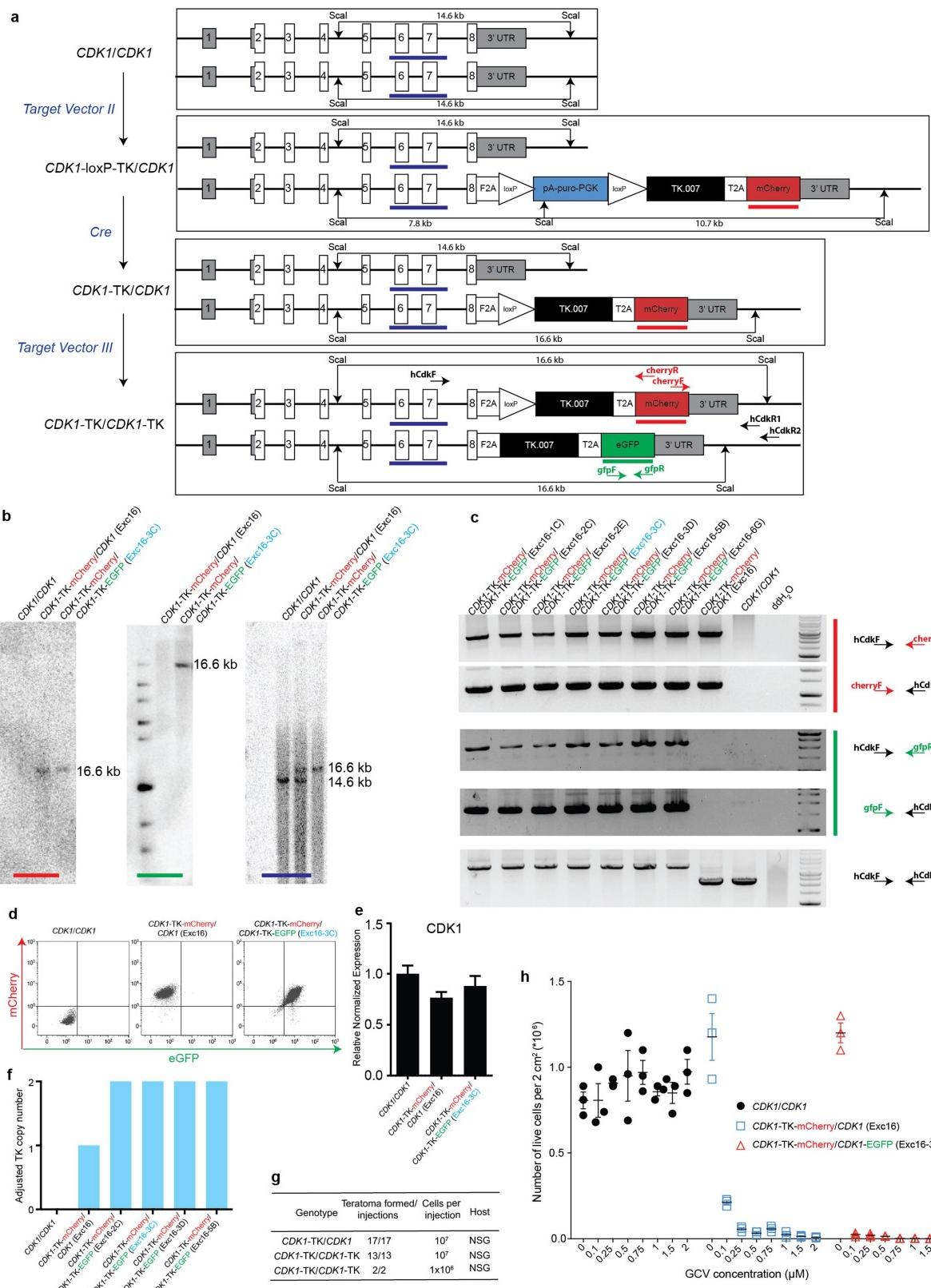
Extended Data Fig. 2 | Generation, genotyping and characterization of human H1 *CDK1–TK/CDK1* and *CDK1–TK/CDK1–TK* ES cells.
a, Generation of human H1 *CDK1–TK/CDK1* and *CDK1–TK/CDK1–TK* ES cells. **b**, Southern blot genotyping of *CDK1–TK/CDK1* clone Exc16, which was used in teratoma assays (Fig. 1c and Extended Data Fig. 7) and *CDK1–TK/CDK1–TK* clone Exc16-3C, which was used in the differentiation assays in Fig. 3. **c**, PCR genotyping of all the correct clones. **d**, Flow cytometry analysis of the *CDK1–TK/CDK1* clone Exc16 and the *CDK1–TK/CDK1–TK* clone Exc16-3C. **e**, SybrGreen qPCR of human

CDK1 expression in H1 wild-type cells, and cells expressing the *CDK1–TK/CDK1* clone Exc16 and the *CDK1–TK/CDK1–TK* clone Exc16-3C. Data are mean ± s.d., $n = 3$. No significant difference was found between groups. **f**, TaqMan qPCR copy-number analysis of TKs of all clones with the correct genotype. **g**, The efficiency of teratoma formation in NSG mice using human H1 ES cells. **h**, Dose–response analysis of wild-type, *CDK1–TK/CDK1* and *CDK1–TK/CDK1–TK* human H1 ES cells. Cells were treated with different GCV concentrations, dissociated and counted after seven days. Data are mean ± s.d., $n = 3$.
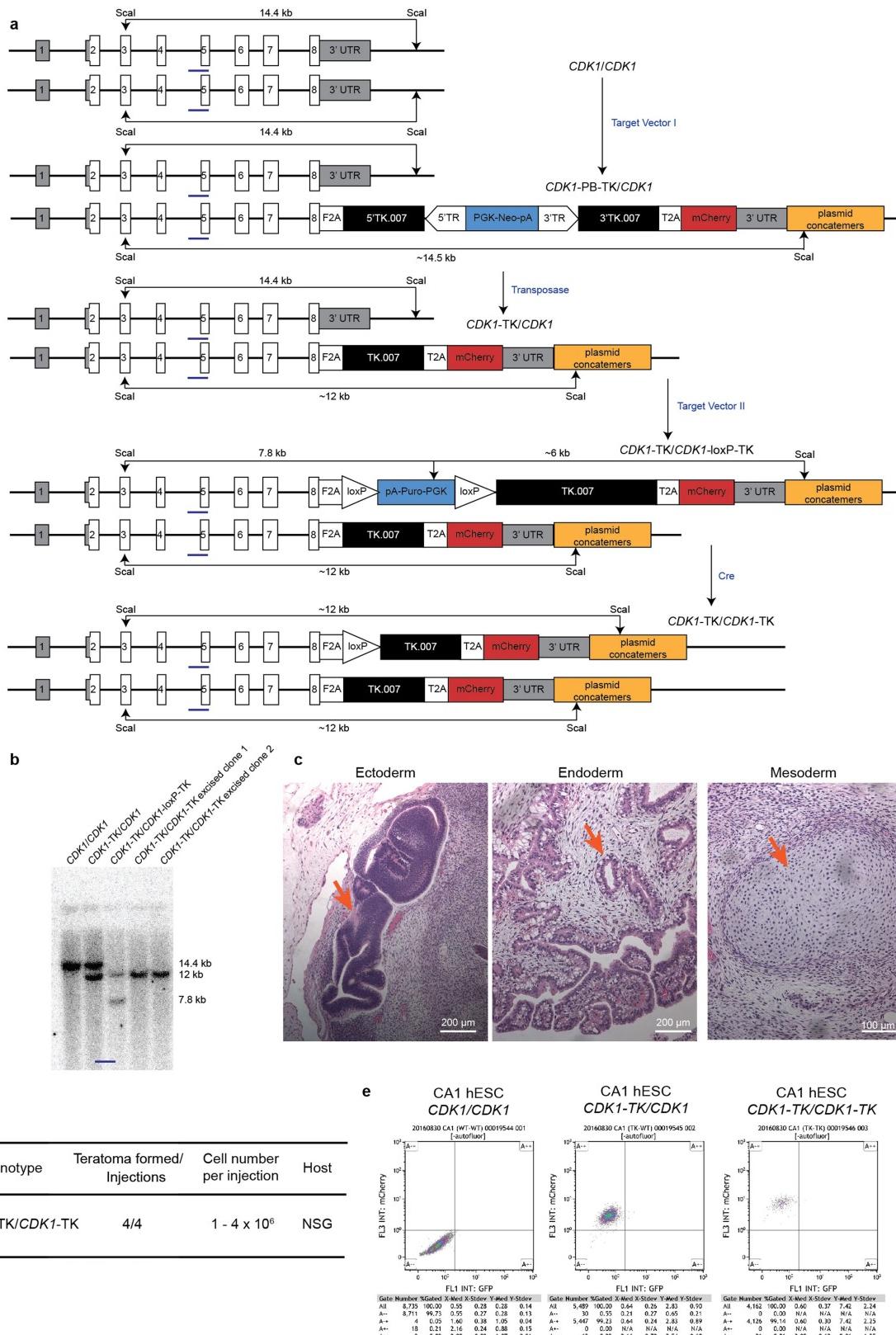
Extended Data Fig. 3 | Generation, genotyping and characterization of human CA1 *CDK1–TK/CDK1* and *CDK1–TK/CDK1–TK* ES cells.
a, Generation of human CA1 *CDK1–TK/CDK1* and *CDK1–TK/CDK1*–TK ES cells. **b**, Southern blot genotyping of human CA1 *CDK1–TK/CDK1* and *CDK1–TK/CDK1–TK* ES cells. The plasmid concatemers are multiple copies of plasmid integration (including backbone). The ampicillin gene in the backbone contains a ScaI restriction enzyme site, which is consistent with the sizes of the band in Southern blots. **c**, Haematoxylin and eosin staining of a *CDK1–TK/CDK1–TK* CA1 ES-cell-derived teratoma. **d**, The efficiency of teratoma formation in NSG mice using human CA1 ES cells. **e**, Flow cytometry analysis shows a direct correlation between the number of *CDK1–TK* alleles and mCherry fluorescence levels.
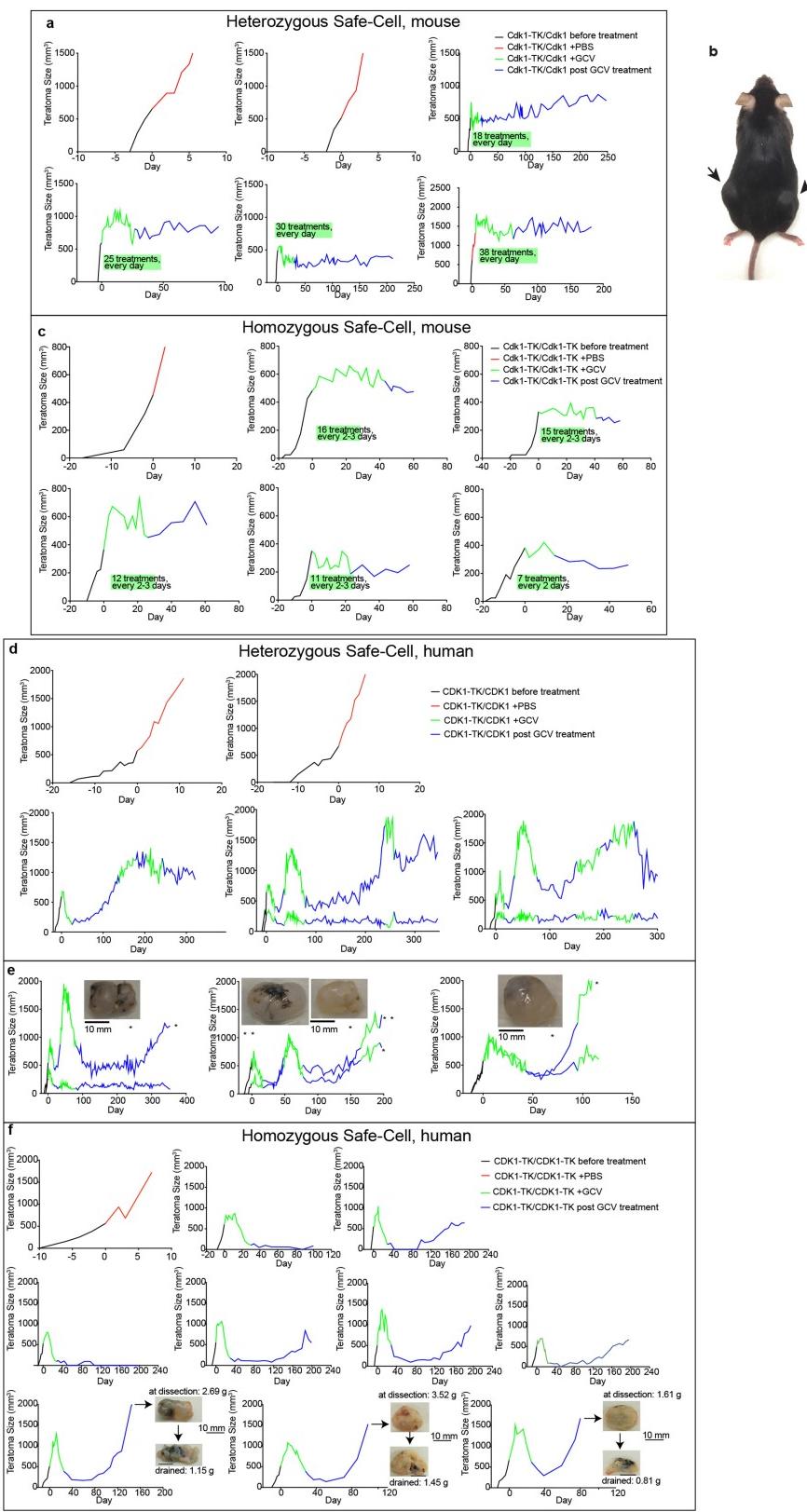
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Growth graphs of mouse and human ES-cell-derived teratomas. a, Growth of teratomas derived from mouse heterozygous safe-cell ES cells (C2 *Cdk1–TK/Cdk1*) **b**, Adult mouse with stabilized subcutaneous tissue (safe-cell ES-cell-derived dormant teratoma), 2.5 months after GCV treatment. **c**, Growth of teratomas derived from mouse homozygous safe-cell ES cells (C2 *Cdk1–TK/Cdk1–TK*). **d**, Growth of teratomas derived from human heterozygous safe-cell ES cells (H1 *CDK1–TK/CDK1*, clone Exc16); daily GCV treatment. **e**, Examples of teratomas from human heterozygous safe-cell ES cells showing cyst formation, images of cystic teratomas at dissection are shown next to the corresponding growth line; daily GCV treatment. The graphs with two lines represent mice that had cells injected into both flanks. The graphs with one line represent mice that had cells injected into one flank. The GCV treatment regime varies among mice because each teratoma behaves differently; we started GCV when the teratoma size started to increase. **f**, Growth of teratomas derived from human homozygous safe-cell ES cells (H1 *CDK1–TK/CDK1–TK*), GCV treatment was every other day. Images of cystic teratomas are shown next to the corresponding growth line, cysts were drained after dissection to show the difference in tumour weight due to the fluid present in the tissue. Each graph represents one mouse. **a**, **c**, **d**, **f**, All replicates of these experiments are shown.
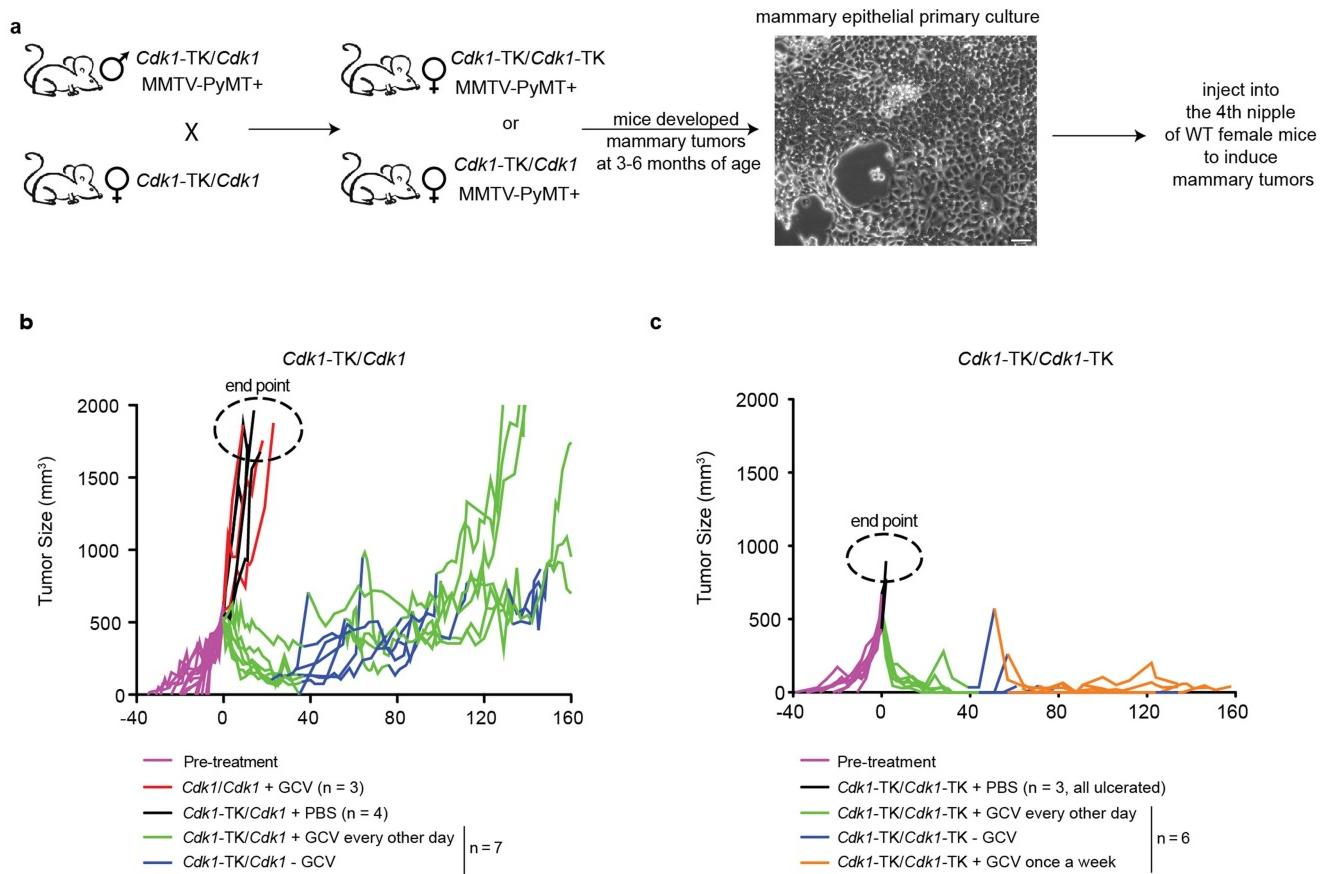
Extended Data Fig. 5 | Breast cancer transplantation assay using heterozygous and homozygous safe-cell mammary tumour cells. **a**, Generation of mouse lines and experimental design. **b**, Growth of mammary gland tumours derived from mouse *Cdk1/Cdk1* and *Cdk1–TK/Cdk1* mammary epithelial cells with PBS or GCV treatment. **c**, Growth of mammary gland tumours derived from mouse *CDK1–TK/CDK1–TK* mammary epithelial cells with PBS or GCV treatment. **b, c,** The sample sizes of each group are indicated at the bottom of the graphs.
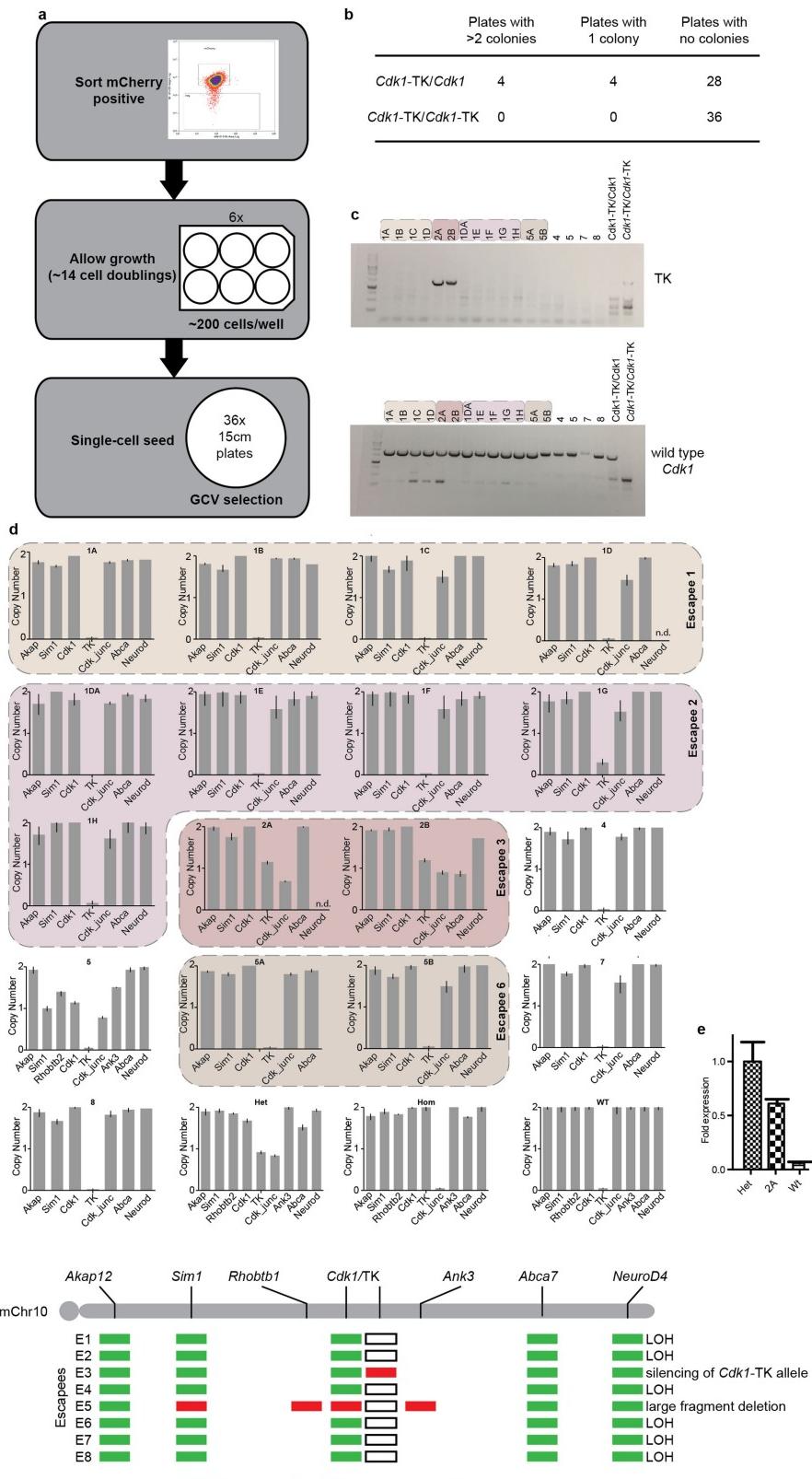
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | In vitro experiments with mouse C2 *Cdk1–TK/Cdk1* **and** *Cdk1–TK/Cdk1–TK* **ES cells and subsequent characterization of escapees. a**, Experimental design: mCherry[+] cells were selected by sorting to ensure that the starting cell population did not contain escapees. These cells were plated on six-well plates (200 cells per well, in a total of 36 wells) and allowed to grow to 14 cell doublings (this was estimated by counting cells in sample wells). The 36 cultures were then resuspended to a single-cell suspension and each was plated in a 15-cm plate (4×10^6 cells). One day after plating, selection with GCV was started and maintained until escapee colonies appeared. **b**, Escapee numbers obtained in 36 independent cultures growing from *Cdk1–TK/Cdk1* and *Cdk1–TK/Cdk1–TK* ES cells. **c**, PCR to determine the presence of TK.

d, TaqMan copy number qPCR analysis of Akap7, *Sim1* and *Cdk1* junction of exon 8 and 3′ UTR, *Neurod*, *Cdk1*, *TK* transgene and *Abca* on mouse chromosome 10. Data are the copy number calculated by CopyCaller Software v.2.1 and the error bars indicate the range from the minimum to the maximum number. $n = 3$. The same colour in the background of **c** and **d** indicates that they are from the same independent culture. n.d., not determined. **e**, qPCR to compare TK expression level in *Cdk1–TK/Cdk1* escapee clone 2A and C2 wild-type ES cells. Data are mean $\pm$ s.e.m., $n = 3$. **f**, Summary of the copy number analysis of mouse *Cdk1–TK/Cdk1* escapees. **a**, **b**, Experiments were repeated twice on a smaller scale but with similar results. **d**, Experiments were repeated twice with similar results.

a

mutation type	consequence	probability designation
type 1	CDEL~~SU~~ → su1 allele	P_1
type 2	~~CDEL~~~~SU~~ → su2 allele	P_2
type 3	dLOH	P_3

b
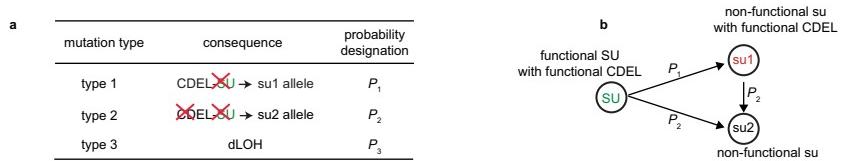
Extended Data Fig. 7 | Schematics of the possible mutation types affecting the CDL–SU allele. a, Three types of mutations that could affect the CDL–SU allele. **b**, Safe-cell allele transition considered in the modelling and Monte Carlo simulations.
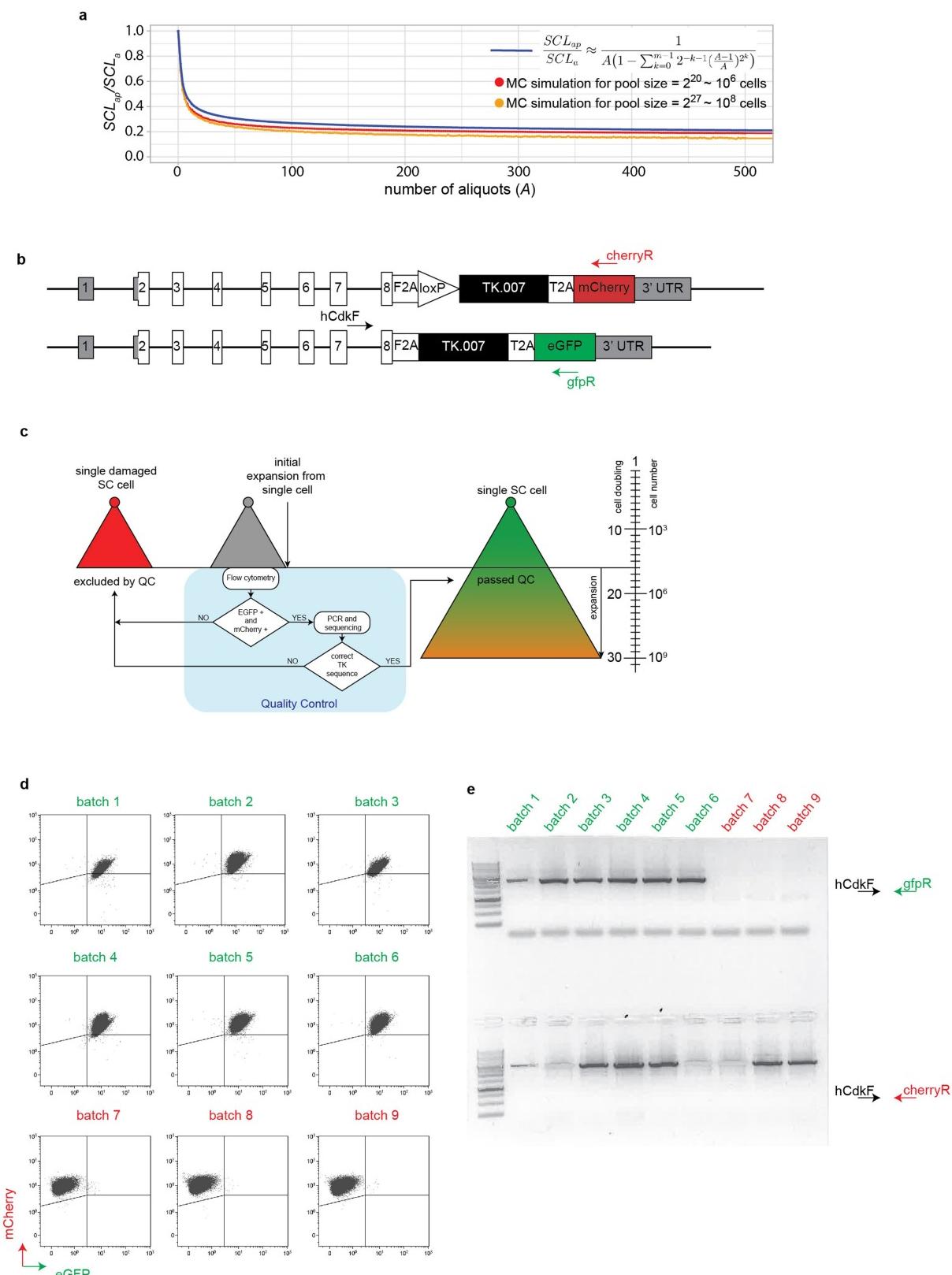
Extended Data Fig. 8 | Quality control of batches generated from single human H1 *CDK1–TK/CDK1–TK* ES cells. a, The drop of SCL due to aliquoting from a pool of cells relative to non-aliquoted batches of the same size. **b**, Schematics of the alleles in the *CDK1–TK/CDK1–TK* human ES cells used in the quality control. **c**, Workflow schematic of performing quality control (QC) on several ES cell batches. **d**, An example of the flow cytometry for the quality control of nine clonally derived batches. **e**, An example of PCR for the quality control of nine clonally derived batches.
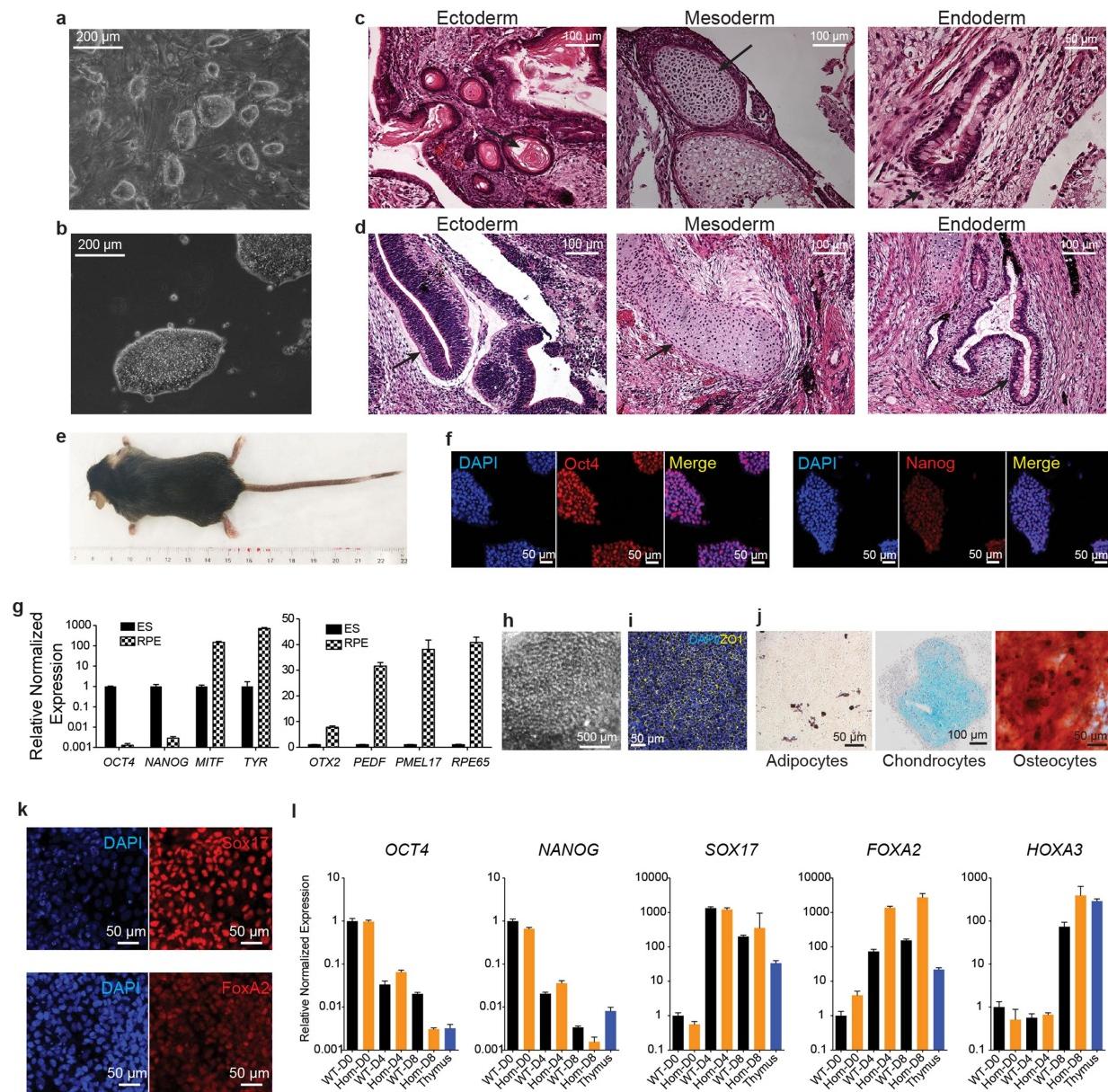
Extended Data Fig. 9 | Mouse and human safe-cell homozygous *CDK1–TK/CDK1–TK* cells demonstrate pluripotency. All experiments were performed using the same clone of mouse C2 or human H1 (Exc16-3C) *CDK1–TK/CDK1–TK* cells. **a**, Bright-field photograph showing mouse homozygous *Cdk1–TK/Cdk1–TK* ES cell morphology. **b**, Bright-field photograph showing human homozygous *CDK1–TK/CDK1–TK* ES cell morphology. **c**, Haematoxylin and eosin staining of a mouse *Cdk1–TK/Cdk1–TK* ES-cell-derived teratoma. **d**, Haematoxylin and eosin staining of a human *CDK1–TK/CDK1–TK* ES-cell-derived teratoma. **e**, An adult *Cdk1–TK/Cdk1–TK* mouse. **f**, OCT4 and NANOG staining of human *CDK1–TK/CDK1–TK* ES cells. **g**, qPCR characterization of

CDK1–TK/CDK1–TK ES cell differentiation into RPE cells. Data are mean ± s.d., $n = 3$. **h**, Bright-field picture of human *CDK1–TK/CDK1–TK* ES-cell-derived RPE cells. **i**, ZO1 staining of human *CDK1–TK/CDK1–TK* ES-cell-derived RPE cells. **j**, Human *CDK1–TK/CDK1–TK* ES-cell-derived adipocytes, chondrocytes and osteocytes. **k**, SOX17 and FOXA2 staining of human *CDK1–TK/CDK1–TK* ES-cell-derived definitive endoderm. **l**, *OCT4* (also known as *POU5F1*), *NANOG*, *SOX17*, *FOXA2* and *HOXA3* qPCR characterization of ES cell (day 0) differentiation into definitive endoderm (day 4) and pharyngeal pouch endoderm (day 8). Data are mean ± s.d., $n = 3$. **a–d**, **f–i**, Experiments were repeated three times with similar results. **j–l**, Experiments were repeated twice with similar results.
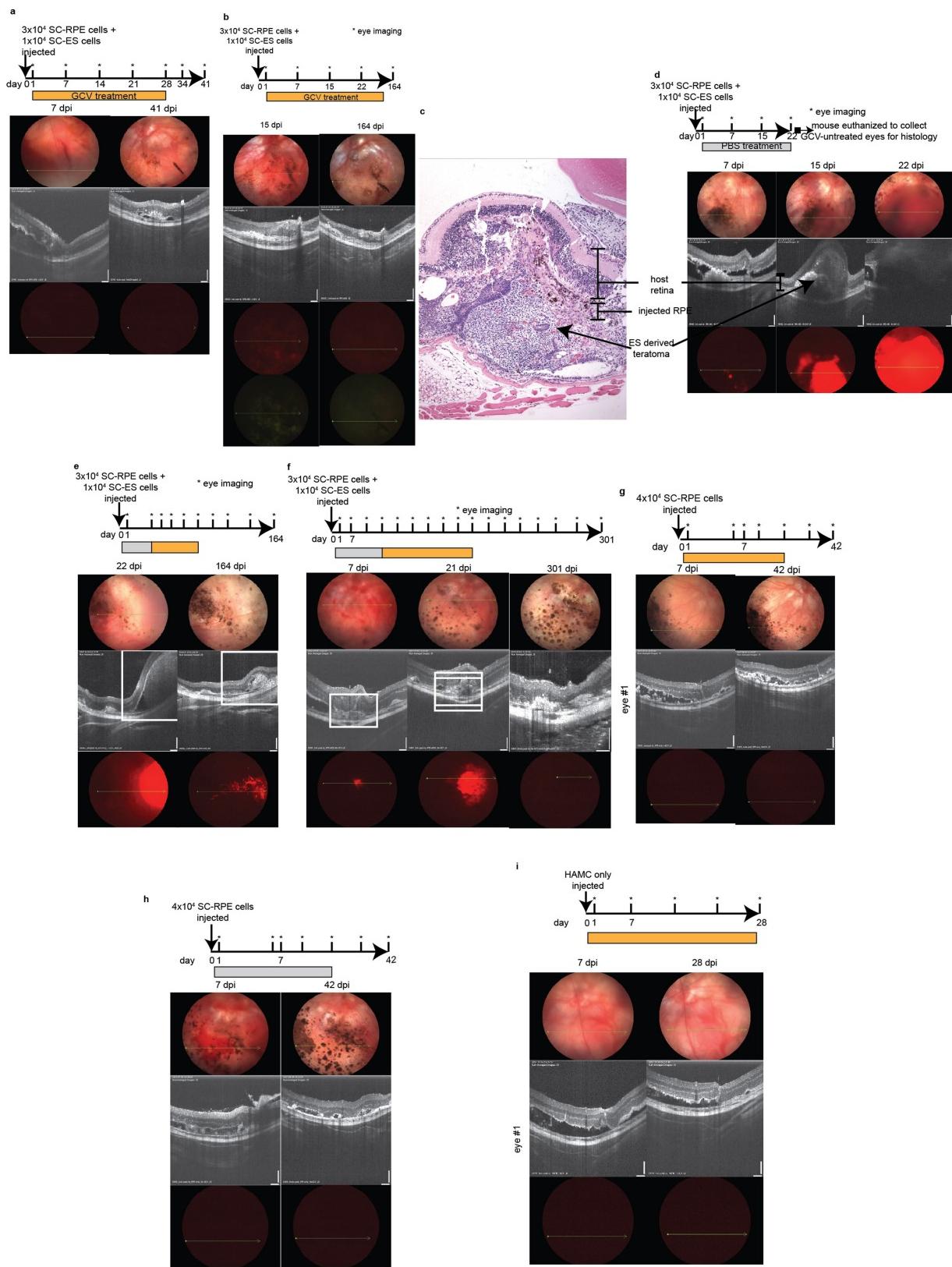
Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Representative images of eyes transplanted with both safe-cell RPE and safe-cell ES cells, and only with safe-cell RPE cells. a, **b**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes transplanted with safe-cell RPE and safe-cell ES cells (four-week GCV treatment). The absence of a mCherry signal indicates that ES cell growth has not occurred. **b**, Bottom, images of the green fluorescence channel are included to illustrate that the observed signal in the red fluorescence channel is actually autofluorescence. This experiment was repeated four times in four mice with similar results. **c**, Histological analysis of the eye presented in **d**. **d**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes transplanted with safe-cell RPE and safe-cell ES cells (PBS treatment). This experiment was repeated twice in two mice with similar results. **e**, **f**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes transplanted with safe-cell RPE and safe-cell ES cells, mCherry signal is detectable and indicates ES cell growth. GCV treatment began three weeks post-injection following an initial PBS treatment. This experiment was repeated four times in three mice with similar results. **g**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes receiving only safe-cell RPE cells (four-week GCV treatment). This demonstrates that GCV treatment did not affect the RPE cells. This experiment was repeated five times in three mice with similar results. **h**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes receiving only safe-cell RPE cells (four-week PBS treatment). This experiment was repeated six times in three mice with similar results. **i**, Fundoscopy, optical coherence tomography and fluorescence imaging of eyes receiving only HAMC (four-week GCV treatment). This experiment was repeated twice in one mouse with similar results.

LETTER

Methicillin–resistant *Staphylococcus aureus* alters cell wall glycosylation to evade immunity

David Gerlach[1,2,13], Yinglan Guo[3,13], Cristina De Castro[4], Sun-Hwa Kim[5], Katja Schlatterer[1,2], Fei-Fei Xu[6], Claney Pereira[6], Peter H. Seeberger[6], Sara Ali[7], Jeroen Codée[7], Wanchat Sirisarn[8], Berit Schulte[2,9], Christiane Wolz[2,9], Jesper Larsen[10], Antonio Molinaro[11], Bok Luel Lee[5], Guoqing Xia[8], Thilo Stehle[3,12,14]* & Andreas Peschel[1,2,14]*

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a frequent cause of difficult-to-treat, often fatal infections in humans[1,2]. Most humans have antibodies against *S. aureus*, but these are highly variable and often not protective in immunocompromised patients[3]. Previous vaccine development programs have not been successful[4]. A large percentage of human antibodies against *S. aureus* target wall teichoic acid (WTA), a ribitol-phosphate (RboP) surface polymer modified with *N*-acetylglucosamine (GlcNAc)[5,6]. It is currently unknown whether the immune evasion capacities of MRSA are due to variation of dominant surface epitopes such as those associated with WTA. Here we show that a considerable proportion of the prominent healthcare-associated and livestock-associated MRSA clones CC5 and CC398, respectively, contain prophages that encode an alternative WTA glycosyltransferase. This enzyme, TarP, transfers GlcNAc to a different hydroxyl group of the WTA RboP than the standard enzyme TarS[7], with important consequences for immune recognition. TarP-glycosylated WTA elicits 7.5–40-fold lower levels of immunoglobulin G in mice than TarS-modified WTA. Consistent with this, human sera contained only low levels of antibodies against TarP-modified WTA. Notably, mice immunized with TarS-modified WTA were not protected against infection with *tarP*-expressing MRSA, indicating that TarP is crucial for the capacity of *S. aureus* to evade host defences. High-resolution structural analyses of TarP bound to WTA components and uridine diphosphate GlcNAc (UDP-GlcNAc) explain the mechanism of altered RboP glycosylation and form a template for targeted inhibition of TarP. Our study reveals an immune evasion strategy of *S. aureus* based on averting the immunogenicity of its dominant glycoantigen WTA. These results will help with the identification of invariant *S. aureus* vaccine antigens and may enable the development of TarP inhibitors as a new strategy for rendering MRSA susceptible to human host defences.

Novel prevention and treatment strategies against major antibiotic-resistant pathogens such as MRSA are urgently needed but are not within reach because some of the most critical virulence strategies of these pathogens are not understood[8]. The pathogenic potential of prominent healthcare-associated (HA)-MRSA and recently emerged livestock-associated (LA)-MRSA strains is thought to rely on particularly effective immune evasion strategies, whereas community-associated (CA)-MRSA strains often produce more aggressive toxins[1,2]. Most humans have high overall levels of antibodies against *S. aureus* as a consequence of preceding infections, but antibody titres differ strongly for specific antigens and are often not protective in immunocompromised patients, for reasons that are not clear[3]. A large percentage of human antibodies against *S. aureus* is directed against WTA[5,9,10],

which is largely invariant. However, some *S. aureus* lineages produce altered WTA, which modulates, for instance, phage susceptibility[7,11].

To investigate whether some prevalent *S. aureus* lineages use additional WTA-targeted strategies to increase their fitness and pathogenicity, we screened *S. aureus* genomes for potential additional paralogues of WTA biosynthesis genes. We found three *S. aureus* prophages that encoded a protein, TarP, that has 27% identity to the WTA-β-GlcNAc
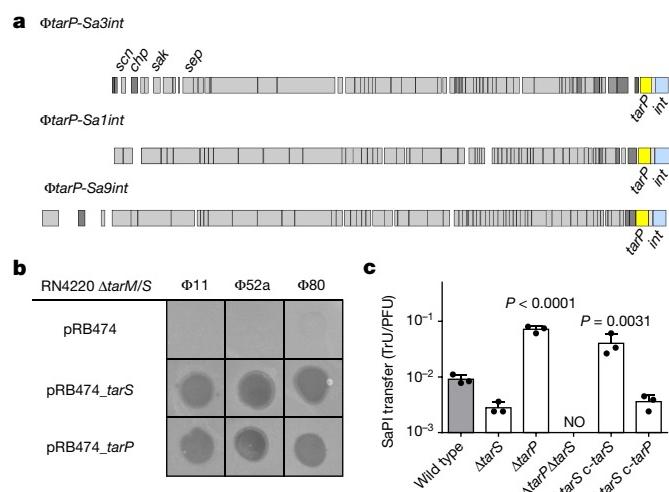
Fig. 1 | The phage-encoded TarP can replace the housekeeping WTA β-GlcNAc transferase TarS. a, TarP is encoded next to different integrase types (*int* gene) in prophages φ*tarP*-Sa3int (with immune evasion cluster *scn*, *chp*, *sak*, *sep*), found in HA-MRSA, and φ*tarP*-Sa1int and φ*tarP*-Sa9int, identified in LA-MRSA. TarP variants in φ*tarP*-Sa1int and φ*tarP*-Sa9int differed from TarP in φ*tarP*-Sa3int in one amino acid each (I8M and D296N, respectively). Both residues are distant from the catalytic centre. **b**, Complementation of *S. aureus* RN4420 Δ*tarM/S* with either *tarS* or *tarP* restores susceptibility to infection by WTA GlcNAc-binding siphophages, as indicated by plaque formation on bacterial lawns. Data shown are representative of three independent experiments. **c**, *tarP* expression reduces siphophage Φ11-mediated transfer of SaPIbov in N315. Values indicate the ratio of transduction units (TrU) to plaque-forming units (PFU) given as mean ± s.d. of three independent experiments. Statistical significances when compared to wild type were calculated by one-way ANOVA with Dunnett's post-test (two-sided) and significant *P* values (*P* ≤ 0.05) are indicated. NO (none obtained) indicates no obtained transductants.

[1]Interfaculty Institute of Microbiology and Infection Medicine, Infection Biology, University of Tübingen, Tübingen, Germany. [2]German Centre for Infection Research (DZIF), Partner Site Tübingen, Tübingen, Germany. [3]Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany. [4]Department of Agricultural Sciences, University of Naples, Naples, Italy. [5]National Research Laboratory of Defense Proteins, College of Pharmacy, Pusan National University, Pusan, South Korea. [6]Max-Planck-Institute for Colloids and Interfaces, Potsdam, Germany. [7]Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands. [8]Lydia Becker Institute of Immunology and Inflammation, Division of Infection, Immunity and Respiratory Medicine, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. [9]Interfaculty Institute of Microbiology and Infection Medicine, Medical Microbiology, University of Tübingen, Tübingen, Germany. [10]Bacteria, Parasites and Fungi, Statens Serum Institut, Copenhagen, Denmark. [11]Department of Chemical Sciences, University of Naples, Naples, Italy. [12]Vanderbilt University School of Medicine, Nashville, TN, USA. [13]These authors contributed equally: David Gerlach, Yinglan Guo. [14]These authors jointly supervised this work: Thilo Stehle, Andreas Peschel. *e-mail: thilo.stehle@uni-tuebingen.de; andreas.peschel@uni-tuebingen.de
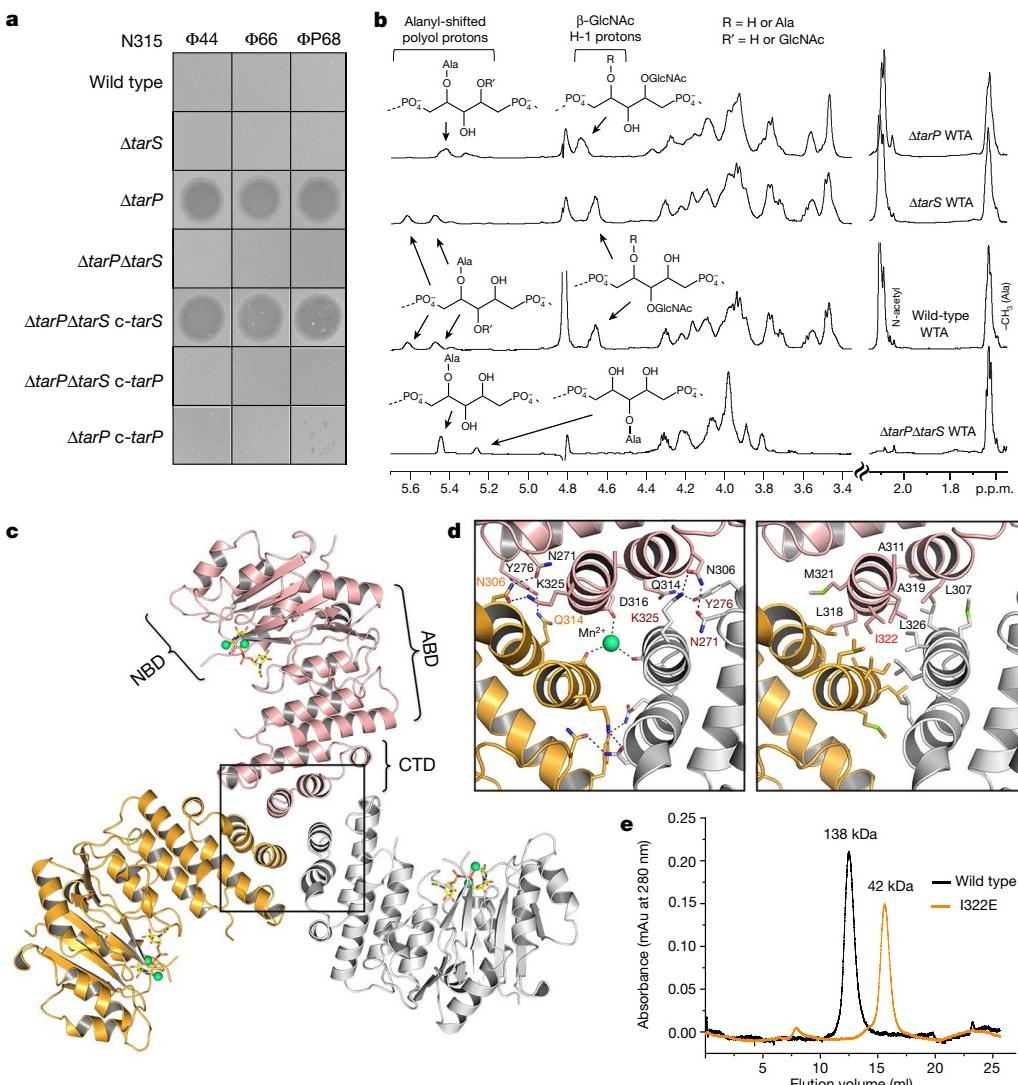
Fig. 2 | TarP protects N315 from podophage infection by alternative glycosylation of WTA at RboP C3. a, Expression of *tarP* renders N315 resistant to podophages. Representative data from three independent experiments are shown. **b**, ¹H NMR spectra reveal different ribitol hydroxyl glycosylation of N315 WTA by TarS (C4) or TarP (C3). The RboP units with attached GlcNAc are depicted above the corresponding proton resonances. Representative data from three experiments are shown. In-depth description of the structural motifs identified in the spectra is given in the Supplementary Information. **c**, Crystal structure of TarP homotrimer (pink, orange, grey) bound to UDP-GlcNAc (yellow) and two Mn²⁺ ions (lime green). The nucleotide-binding domain (NBD), acceptor-

binding domain (ABD), and C-terminal trimerization domain (CTD) of the pink monomer are labelled. **d**, Views into the trimer interface (boxed in **c**). Left, polar interactions. Hydrogen bonds and salt bridges are shown as black dashed lines. The Mn²⁺ is 2.1 Å from each Asp316 carboxylate. Right, hydrophobic interactions, with the mutated residue Ile322 highlighted in red. **e**, Size-exclusion chromatography elution profiles. Based on calibration of the column, the TarP wild-type and I322E mutant proteins have estimated molecular weights of 138 kDa ($n = 8$) and 42 kDa ($n = 3$), respectively, in agreement with the calculated molecular weights of 120 kDa for a TarP trimer and 40 kDa for monomeric TarP.

transferase TarS[7] (Fig. 1a). *tarP* was found exclusively in isolates of the prominent HA-MRSA CC5[12], on a prophage that also encoded the *scn*, *chp* and *sak* immune evasion genes[13], and on two other prophages in the emerging LA-MRSAs CC398[14] and CC5[15]. All *tarP*-harbouring genomes also contained *tarS*.

When *tarP* from CC5 HA-MRSA strain N315 was expressed in a WTA glycosylation-deficient mutant of laboratory strain RN4220[7], it restored WTA glycosylation (Extended Data Fig. 1a) and susceptibility to siphophages, which need RboP WTA GlcNAc as a binding motif[16] (Fig. 1b). The presence of β-GlcNAc on WTA is essential for full β-lactam resistance in MRSA strains[7]. When *tarP* was expressed in a WTA glycosylation-deficient mutant of CA-MRSA strain MW2 (CC1), it restored full oxacillin resistance (Extended Data Fig. 1b), confirming that *tarP* can replace *tarS* in several key interactions.

The expression of TarP led to susceptibility to siphophages, albeit to a lower extent than TarS (Extended Data Fig. 1c), although TarP did not

incorporate less GlcNAc into WTA than TarS (Extended Data Fig. 1d, Supplementary Table 3). Similarly, the siphophage-mediated horizontal transfer of an *S. aureus* pathogenicity island was reduced about tenfold in *S. aureus* N315 expressing *tarP*, compared to the same strain expressing only *tarS* (Fig. 1c), suggesting that TarP and TarS glycosylate WTA differently. Notably, N315 was resistant to podophages, but inactivation of *tarP* (but not of *tarS*) rendered it susceptible to podophages (Fig. 2a). We analysed the overall effect of *tarP* on podophage susceptibility patterns in 90 clinical CC5 and CC398 isolates and found that none of the *tarP*-containing strains, but all of the *tarP*-lacking strains, were susceptible to podophages (Extended Data Table 1). Thus, TarP causes podophage resistance and TarP-mediated modification of WTA is distinct from that mediated by TarS. Nuclear magnetic resonance (NMR) analyses revealed that both TarP and TarS add GlcNAc to WTA in the β-configuration. However, the attachment site in RboP differs: TarS glycosylates the C4 position[17] whereas TarP attaches GlcNAc
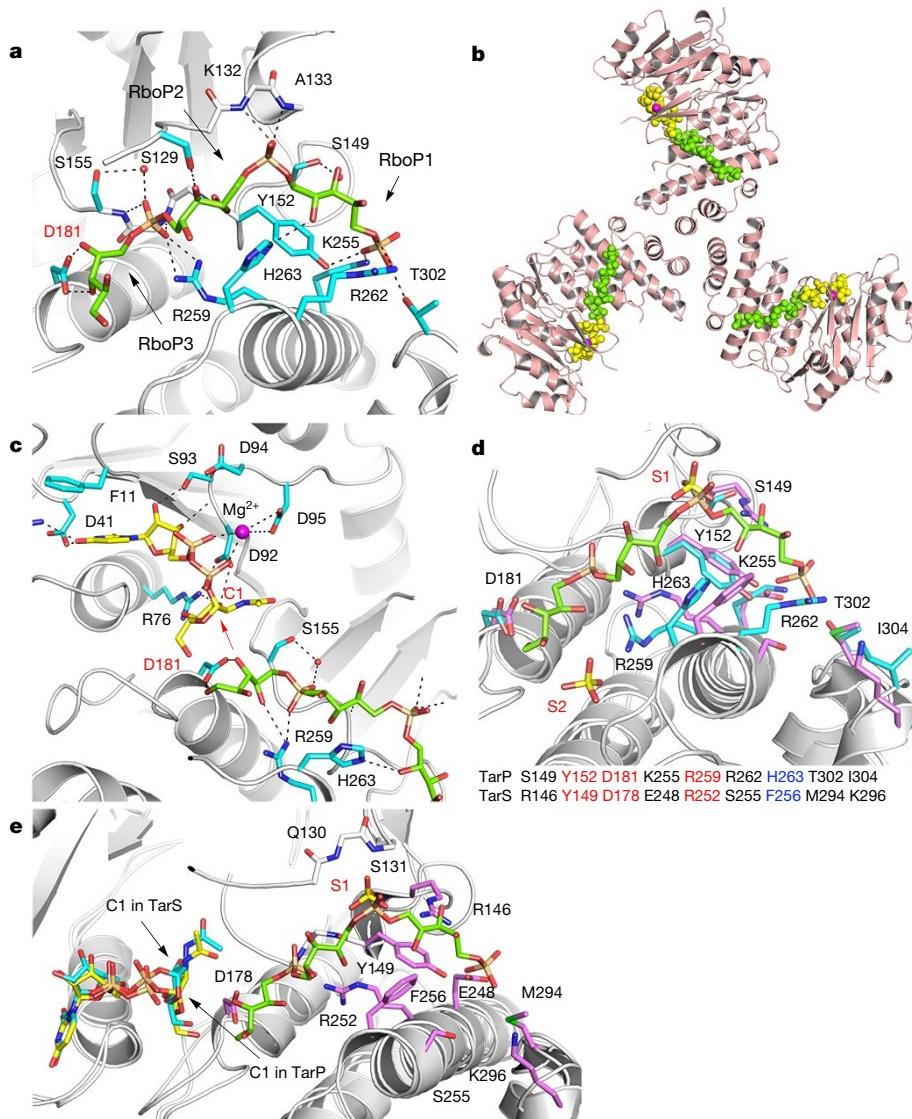
Fig. 3 | Interactions of TarP with UDP-GlcNAc and D-ribitol-5-phosphate trimer (3RboP), and comparison of polyRboP binding sites of TarP and TarS. a, 3RboP binding site in the TarP–3RboP complex, with key amino acids shown (cyan). Asp181 is highlighted in red. The ribitol of 3RboP is coloured green and D-ribitol-5-phosphate units 1, 2 and 3 (RboP1, RboP2, and RboP3) are labelled. Hydrogen bonds and salt bridges are shown as black dashed lines. **b**, Ternary complex of TarP with UDP-GlcNAc and 3RboP. UDP-GlcNAc, Mg^{2+} and 3RboP are shown as full-atom models coloured yellow, magenta, and green, respectively. **c**, View into the active site of TarP. C1 of UDP-GlcNAc and Asp181 are highlighted with red labels. The arrow indicates how the C3-hydroxyl in RboP3 could nucleophilically attack GlcNAc C1. **d**, Comparison of the polyRboP-binding site of TarP with the corresponding region in TarS. Residues of TarP and 3RboP are coloured as in **a**. TarS residues are coloured violet and the two sulfates are labelled S1 and S2. Only residues of TarP are labelled, for clarity. Key TarP and TarS residues lining the polyRboP-binding site are shown at the bottom, with three identical (red) and one conserved amino acids (blue). **e**, Superposition of UDP-GlcNAc-bound TarS with the ternary TarP–UDP-GlcNAc–3RboP complex. UDP-GlcNAc and 3RboP bound to TarP are coloured as in **b**, whereas UDP-GlcNAc bound to TarS is coloured in cyan. Only the TarS residues are shown (coloured as in **d**), for clarity. The arrows indicate the C1 positions of UDP-GlcNAc in TarP and TarS.

to C3 (Fig. 2b, Extended Data Fig. 2, Supplementary Table 2). This difference may be crucial for impairing phage infection. Moreover, NMR analysis revealed that TarP is dominant over TarS because in N315, which bears both genes, GlcNAc was almost exclusively attached to RboP C3 (Fig. 2b).

We solved the TarP structure at high resolution to elucidate how TarP generates a different glycosylation product from TarS. Like TarS[18], TarP forms stable homotrimers, but it uses a different trimerization strategy because it lacks the C-terminal trimerization domain found in TarS (Fig. 2c, Extended Data Fig. 3). Instead, hydrophobic and polar interactions of a small helical C-terminal domain generate the TarP trimer (Fig. 2d, e). WTA polymers comprising three or six RboP repeating units (3RboP or 6RboP-$(CH_2)_6NH_2$, respectively) were synthesized and used for soaking TarP crystals (Supplementary Information Fig. 2, 3), yielding the first protein structure visualizing the binding of a WTA-based polymer (Fig. 3, Extended Data Fig. 4). In the ternary complex TarP–UDP-GlcNAc–3RboP, the distance between the C3-hydroxyl of the third unit of 3RboP (RboP3) and the anomeric C1 of GlcNAc is 4.2 Å. Furthermore, at 3.1 Å, Asp181 is well within hydrogen bonding distance of the C3-hydroxyl of RboP3. The observed distances and geometry nicely explain the unusual glycosylation of WTA at the C3-hydroxyl. We propose that TarP uses a direct S_N2-like glycosyltransferase reaction, as discussed for other inverting GT-A fold enzymes[19,20]. In this mechanism, Asp181 would act as the catalytic base, deprotonating the C3-hydroxyl on RboP3 and enabling a nucleophilic attack on

the GlcNAc C1, thus yielding a β-O-GlcNAcylated polyRboP (Fig. 3c). Mutagenesis of Asp181 to alanine rendered TarP inactive, supporting this putative mechanism (Extended Data Table 2).

The ternary structure of TarP–UDP-GlcNAc–3RboP allows us to predict how polyRboP binds to the homologous TarS enzyme. Three residues that are critical for binding and catalysis (including Asp181) are identical in TarP and TarS, while five other residues differ[18] (Fig. 3d). Lys255 and Arg262, for instance, which interact electrostatically with a WTA phosphate group in TarP, are replaced with Glu248 and Ser255, respectively, in TarS, which may lead to reduced affinity for WTA and might explain why TarP is dominant over TarS in vivo. On the basis of the location of UDP-GlcNAc, the identical Tyr149, Asp178 and Arg252 side chains, the conserved aromatic side chain of Phe256, and a site that contains a bound sulfate ion from the crystallization solution (S1) and probably binds phosphate in TarS (Fig. 3e), the polyRboP chain would be shifted to the upper right, and the relative position of RboP units in the binding site would be altered in TarS. Such an altered binding mode would move the C4-hydroxyl of the target RboP towards C1 of GlcNAc, allowing TarS to glycosylate at the C4 position.

S. aureus WTA is a dominant antigen for adaptive immune responses[5,9]. The observation that the position of GlcNAc on RboP had a profound impact on binding by podophage receptors raises the question of whether human antibodies also discriminate between the two isomeric polymers and whether MRSA clones use TarP to subvert immune recognition. We analysed several human antibody
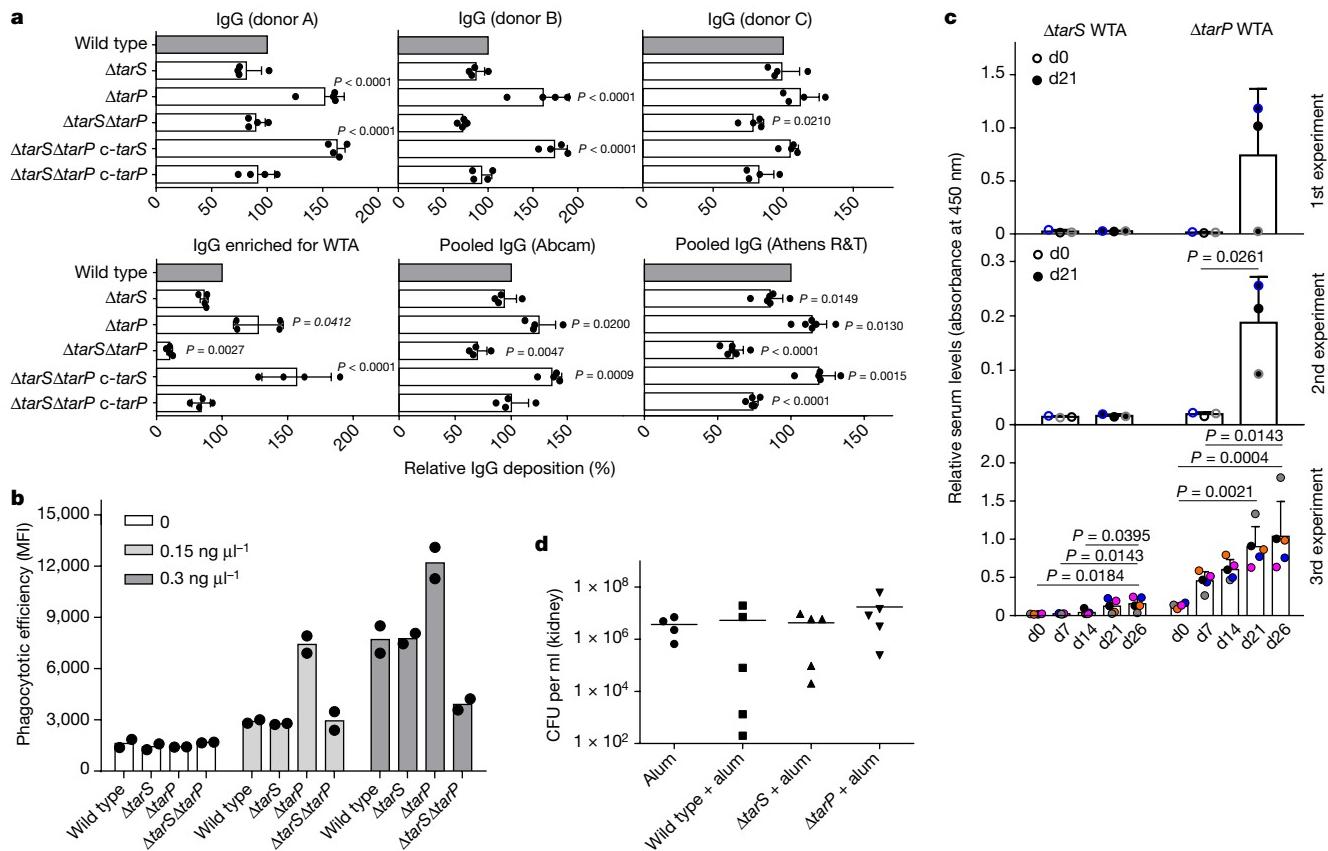
Fig. 4 | TarP attenuates immunogenicity of WTA. a, TarP expression reduces deposition of IgG from human serum on N315 cells. The protein A gene *spa* was deleted in all strains. Top, human IgG isolated from three individual healthy donors (A, B, and C; $n = 4$); bottom, left, IgG from human serum enriched for RN4220 WTA binding ($n = 4$); middle and right, pooled human IgG from different suppliers (Abcam, $n = 4$; Athens R&T, $n = 6$). Results were normalized against wild type and shown as means with s.d. of n experiments. *P* values for comparison with wild type were calculated by one-way ANOVA with Dunnett's post-test (two-sided), and $P \leq 0.05$ was considered significant. Significant *P* values are displayed. **b**, TarP reduces neutrophil phagocytosis of N315 strains lacking protein A, opsonized with indicated concentrations of IgG enriched for WTA binding. Values are depicted as mean fluorescence intensity (MFI). Means of two dependent replicates of a representative experiment are shown.

The other two representative experiments can be found in Extended Data Fig. 1g. **c**, TarP abrogates IgG response of mice towards WTA. For each experiment, WTA from N315 Δ*tarP* or Δ*tarS* was isolated independently. At least three mice per group were vaccinated and analysed for specific IgG at indicated time points after vaccination. Results are depicted as mean absorbance with s.d. Individual mice are indicated by colour. Increase in IgG levels was assessed by one-way ANOVA with Tukey's post-test (two-sided). Significant differences ($P \leq 0.05$) are indicated with corresponding *P* values. **d**, Vaccination with WTA does not protect mice against *tarP*-expressing N315, as shown for bacterial loads in kidney upon intravenous infection. No significant differences between groups of either five vaccinated mice or four mice for the alum control group (means indicated), calculated by one-way ANOVA, were observed.

preparations for their capacity to opsonize a panel of N315 strains with or without *tarP* and/or *tarS*. The mutant lacking any WTA glycosylation bound the lowest amount of IgG compared to WTA glycosylation-positive strains (Fig. 4a), demonstrating that glycosylated WTA is a prominent *S. aureus* antigen in humans. Exclusive expression of *tarS* led to strongly increased IgG binding compared to the glycosylation-deficient mutant, indicating that β-GlcNAc on RboP C4 is an important epitope for human anti-*S. aureus* antibodies. By contrast, expression of *tarP* in the presence or absence of *tarS* led to only slightly increased IgG binding compared to the glycosylation-deficient mutant. The capacity of TarP to impair the deposition of IgG on *S. aureus* differed with individual serum donors and reached average levels in pooled serum preparations (Fig. 4a). When *tarP* was deleted in three further CC5 isolates, they showed similarly increased capacities to bind human serum antibodies compared to the wild-type strains (Extended Data Fig. 1e). Additionally, *tarP* deletion led to a substantially increased capacity of human neutrophils to phagocytose opsonized *S. aureus* (Fig. 4b, Extended Data Fig. 1g). Thus, only a small percentage of *S. aureus*-specific antibodies can bind WTA with β-GlcNAc on RboP C3, and *tarP*-expressing *S. aureus* are less likely to be detected and eliminated by human phagocytes.

We purified N315 WTA that had been glycosylated by TarS or TarP and used it to immunize mice. Antibodies binding to regular (TarS-modified) WTA increased continuously over three weeks after vaccination (Fig. 4c). By contrast, no or only very low amounts of IgG directed against TarP-glycosylated WTA emerged, indicating that WTA modified at RboP C3 is much less immunogenic than WTA modified at RboP C4. This experiment was repeated three times with three different WTA preparations and yielded broadly similar data.

Vaccination with *S. aureus* WTA bearing GlcNAc at RboP C4 protects mice against infection by CA-MRSA strains USA300 (CC8) or USA400 (CC1), which both lack *tarP*[5,21]. Remarkably, vaccination with regular (TarS-modified) or TarP-modified WTA did not lead to any notable protection against subsequent infection with *tarP*-expressing N315 compared to mock vaccination, despite the robust antibody response against regular WTA (Fig. 4d). Together, our results demonstrate that *tarP* protects *S. aureus* against adaptive host defences by allowing bacteria to evade recognition by preexisting anti-*S. aureus* antibodies and by exploiting the poor immunogenicity of TarP-modified WTA.

It is possible that TarP-modified WTA mimics a currently unknown autoantigen and is therefore hardly immunogenic. On the other hand, regular *S. aureus* WTA can be ingested by antigen-presenting cells

and presented to T cells, in a largely unexplored way, thereby evoking specific immunoglobulins and immunological memory[22,23]. It is possible that TarP-modified WTA is refractory to this process. Thus, TarS- and TarP-modified WTA could be helpful for decoding glycopolymer presentation pathways and for defining the most promising WTA epitopes for the development of protective vaccines against *S. aureus*.

Protection against *S. aureus* infections is urgently needed, in particular for hospitalized and immunocompromised patients[2,4]. Antibodies can in principle protect against *S. aureus*, but their titres and specificities vary largely among humans and they are often not protective in immunocompromised patients[3], probably in particular against *S. aureus* clones that mask dominant epitopes, for instance using TarP. Unfortunately, all previous human vaccination attempts with protein or glycopolymer antigens have failed, for reasons that are unclear[24]. Our study identifies a new strategy used by pandemic MRSA clones to subvert antibody-mediated immunity, which should be considered in future vaccination approaches. *S. aureus* WTA with GlcNAc at RboP C3 has been reported as a type-336 antigen, but was not further explored[25]. We found that *tarP* is present in type-336 *S. aureus* (Extended Data Fig. 1f). However, TarP-modified WTA is a very poor antigen and vaccines directed against GlcNAc at WTA RboP C3 or C4 may fail against many of the pandemic MRSA clones. The structural characterization of TarP will instruct the development of specific TarP inhibitors that could become important in combination with anti-WTA vaccines or antibiotic therapies. We found *tarP*-encoding prophages in 70–80% of south-west German HA-MRSA CC5 and 40% of Danish LA-MRSA CC398 isolates (Extended Data Table 1), pointing to a crucial role of *tarP* in the fitness of these lineages and raising concerns of further dissemination by horizontal gene transfer. TarP is a new and probably crucial component of the *S. aureus* virulence factor arsenal[26,27], highlighting the important roles of adaptive immunity and its evasion in *S. aureus* infections.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0730-x.

1. Tong, S. Y., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G. Jr. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin. Microbiol. Rev.* **28**, 603–661 (2015).
2. Lee, A. S. et al. Methicillin-resistant *Staphylococcus aureus*. *Nat. Rev. Dis. Primers* **4**, 18033 (2018).
3. Stentzel, S. et al. Specific serum IgG at diagnosis of *Staphylococcus aureus* bloodstream invasion is correlated with disease progression. *J. Proteomics* **128**, 1–7 (2015).
4. Missiakas, D. & Schneewind, O. *Staphylococcus aureus* vaccines: deviating from the carol. *J. Exp. Med.* **213**, 1645–1653 (2016).
5. Lehar, S. M. et al. Novel antibody–antibiotic conjugate eliminates intracellular *S. aureus*. *Nature* **527**, 323–328 (2015).
6. Weidenmaier, C. & Peschel, A. Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. *Nat. Rev. Microbiol.* **6**, 276–287 (2008).
7. Brown, S. et al. Methicillin resistance in *Staphylococcus aureus* requires glycosylated wall teichoic acids. *Proc. Natl Acad. Sci. USA* **109**, 18909–18914 (2012).
8. Tacconelli, E. et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).
9. Kurokawa, K. et al. Glycoepitopes of staphylococcal wall teichoic acid govern complement-mediated opsonophagocytosis via human serum antibody and mannose-binding lectin. *J. Biol. Chem.* **288**, 30956–30968 (2013).
10. Lee, J. H. et al. Surface glycopolymers are crucial for in vitro anti-wall teichoic acid IgG-mediated complement activation and opsonophagocytosis of *Staphylococcus aureus*. *Infect. Immun.* **83**, 4247–4255 (2015).
11. Winstel, V. et al. Wall teichoic acid structure governs horizontal gene transfer between major bacterial pathogens. *Nat. Commun.* **4**, 2345 (2013).
12. Nübel, U. et al. Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *Proc. Natl Acad. Sci. USA* **105**, 14130–14135 (2008).
13. McCarthy, A. J. & Lindsay, J. A. *Staphylococcus aureus* innate immune evasion is lineage-specific: a bioinfomatics study. *Infect. Genet. Evol.* **19**, 7–14 (2013).
14. Bal, A. M. et al. Genomic insights into the emergence and spread of international clones of healthcare-, community- and livestock-associated meticillin-resistant *Staphylococcus aureus*: Blurring of the traditional definitions. *J. Glob. Antimicrob. Resist.* **6**, 95–101 (2016).
15. Hau, S. J., Bayles, D. O., Alt, D. P., Frana, T. S. & Nicholson, T. L. Draft genome sequences of 63 swine-associated methicillin-resistant *Staphylococcus aureus* sequence type 5 isolates from the United States. *Genome Announc.* **5**, e01081-17 (2017).
16. Xia, G. et al. Wall teichoic acid-dependent adsorption of staphylococcal siphovirus and myovirus. *J. Bacteriol.* **193**, 4006–4009 (2011).
17. Vinogradov, E., Sadovskaya, I., Li, J. & Jabbouri, S. Structural elucidation of the extracellular and cell-wall teichoic acids of *Staphylococcus aureus* MN8m, a biofilm forming strain. *Carbohydr. Res.* **341**, 738–743 (2006).
18. Sobhanifar, S. et al. Structure and mechanism of *Staphylococcus aureus* TarS, the wall teichoic acid β-glycosyltransferase involved in methicillin resistance. *PLoS Pathog.* **12**, e1006067 (2016).
19. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
20. Kozmon, S. & Tvaroska, I. Catalytic mechanism of glycosyltransferases: hybrid quantum mechanical/molecular mechanical study of the inverting *N*-acetylglucosaminyltransferase I. *J. Am. Chem. Soc.* **128**, 16921–16927 (2006).
21. Takahashi, K. et al. Intradermal immunization with wall teichoic acid (WTA) elicits and augments an anti-WTA IgG response that protects mice from methicillin-resistant *Staphylococcus aureus* infection independent of mannose-binding lectin status. *PLoS One* **8**, e69739 (2013).
22. Weidenmaier, C., McLoughlin, R. M. & Lee, J. C. The zwitterionic cell wall teichoic acid of *Staphylococcus aureus* provokes skin abscesses in mice by a novel CD4+ T-cell-dependent mechanism. *PLoS One* **5**, e13227 (2010).
23. Wanner, S. et al. Wall teichoic acids mediate increased virulence in *Staphylococcus aureus*. *Nat. Microbiol.* **2**, 16257 (2017).
24. Pozzi, C. et al. Vaccines for *Staphylococcus aureus* and target populations. *Curr. Top. Microbiol. Immunol.* **409**, 491–528 (2017).
25. Fattom, A., Sarwar, J., Kossaczka, Z., Taylor, K. & Ennifar, S. Method of protecting against staphylococcal infection. US Patent US20060228368A1 (2006).
26. Thammavongsa, V., Kim, H. K., Missiakas, D. & Schneewind, O. Staphylococcal manipulation of host immune responses. *Nat. Rev. Microbiol.* **13**, 529–543 (2015).
27. Spaan, A. N., Surewaard, B. G., Nijland, R. & van Strijp, J. A. Neutrophils versus *Staphylococcus aureus*: a biological tug of war. *Annu. Rev. Microbiol.* **67**, 629–650 (2013).

Author contributions D.G. characterized TarP in vivo and its genomic context, created mutants, designed experiments, purified WTA, and performed experiments with human IgGs. Y.G. designed experiments, purified proteins, crystallized proteins, solved the structures, and performed in vitro analysis of TarP. C.D.C. performed NMR experiments. C.D.C. and A.M. analysed the NMR data and wrote the NMR discussion. S.-H.K. performed and B.L.L. designed and interpreted mouse immunization and infection experiments. K.S. designed IgG deposition experiments. B.S. and C.W. collected and characterized CC5 MRSA strains. J.L. collected and characterized CC398 strains. J.L. and C.W. analysed *S. aureus* genomes. F.-F.X., C.P., and P.H.S. designed and synthesized 3RboP. S.A. and J.C. designed and synthesized 6RboP-$(CH_2)_6NH_2$. W.S. performed MIC experiments. G.X. identified *tarP*, and characterized and interpreted MIC data. D.G., Y.G., A.P., T.S., and G.X. designed the study, analysed results, and wrote the paper.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Bacterial strains and growth conditions. *S. aureus* strains N315, RN4220, and MW2 (wild type and mutants) were used for this study. Collections of CC5 isolates of the Rhine-Hesse pulsed-field gel electrophoresis type[28] and of the LA-MRSA lineage CC398 from the Danish Statens Serum Institut[29,30] were analysed for the presence of *tarP* and for podophage susceptibility. Additionally, 48 *spa*-type t002 (ST5) and 16 *spa*-type t003 (ST225) isolates were obtained from the MRSA collection of the University Hospital Tübingen and analysed for *tarP* presence by PCR. *S. aureus* strains were cultivated in tryptic soy broth (TSB) or basic medium (BM; 1% tryptone, 0.5% yeast extract, 0.5% NaCl, 0.1% glucose, 0.1% K_2HPO_4, w/v). MICs of oxacillin were determined by microbroth dilution according to established guidelines[31].

Experiments with phages. *tarP*-encoding phages were identified in genome sequences using the webtool Phaster[32] in representative strains listed with GenBank accession: Φ*tarP*-Sa3int with immune evasion cluster (IEC) in CC5 (strain N315, BA000018.3), Φ*tarP*-Sa1int, found in LA-MRSA of CC5 (strain ISU935, CP017090), and Φ*tarP*-Sa9int found in CC398 (strain E154, CP013218).

Phage susceptibility was determined using a soft-agar overlay method[16]. In brief, 10 μl phage lysate of 10^4–10^6 PFU was dropped onto soft agar containing 100 μl bacterial suspension (OD_{600} of 0.1). Plates were incubated at 37 °C overnight. The efficiency of plating was determined as described[33]. Transfer of SaPIs was determined according to previously described methods[11]. In brief, SaPI particle lysates were generated from *S. aureus* strain JP1794, which encodes a SaPI with a resistance marker for tetracycline[34]. PFU of SaPI lysate was determined on RN4220. 200 μl bacterial culture (OD_{600} of 0.5) was mixed with 100 μl of SaPI particle lysate (SaPIbov1 (Φ11), 10^6 PFU/ml), incubated at 37 °C for 15 min. Appropriate dilutions were plated on TSB plates containing 3 μg/ml of tetracycline, and CFU were checked after overnight incubation.

WTA isolation and structure analysis. *WTA from S. aureus* was isolated and purified according to previously described methods[11]. In brief, WTA was released from purified peptidoglycan by treatment with 5% trichloroacetic acid and dialysed extensively against water using a Spectra/Por3 dialysis membrane (MWCO of 3.5 kDa; VWR International GmbH). Obtained soluble WTA was quantified by determining the content of phosphate[35] and GlcNAc[36]. For PAGE analysis of WTA, samples (400 nmol of phosphate per lane) were applied to a 26% polyacrylamide (Rotiphorese Gel 40 (19:1)) resolving gel and separated at 25 mA for 16 h[37]. The gel was equilibrated in a solution of 40% ethanol and 5% acidic acid at room temperature for 1 h and the WTA ladders were visualized by incubation with alcian blue (0.005%) for several hours.

NMR spectroscopy experiments were carried out on a Bruker DRX-600 spectrometer equipped with a cryo-probe, at 288 K (WT-WTA, TarS-WTA, and TarP-WTA) or 298 K (double-mutant WTA lacking any glycosylation). Chemical shifts of spectra recorded in D_2O were calculated in p.p.m. relative to internal acetone (2.225 and 31.45 p.p.m.). The spectral width was set to 10 p.p.m. and the frequency carrier placed at the residual HOD peak, suppressed by pre-saturation. Two-dimensional spectra (TOCSY, gHSQC, gHMBC, and HSQC-TOCSY) were measured using standard Bruker software. For all experiments, 512 FIDs of 2,048 complex data points were collected, 32 scans per FID were acquired for homonuclear spectra, and 20 or 100 ms of mixing time was used for TOCSY spectra. Heteronuclear ^{1}H-^{13}C spectra were measured in the ^{1}H-detected mode, gHSQC spectrum was acquired with 40 scans per FID, the GARP sequence was used for ^{13}C decoupling during acquisition; gHMBC scans doubled those of gHSQC spectrum. As for HSQC-TOCSY, the multiplicity editing during selection step version was used, scans tripled those of the HSQC spectrum and two experiments were acquired by setting the mixing time to 20 or 80 ms. During processing, each data matrix was zero-filled in both dimensions to give a matrix of $4K \times 2K$ points and was resolution-enhanced in both dimensions by a cosine-bell function before Fourier transformation; data processing and analysis were performed with the Bruker Topspin 3 program.

Molecular biology. All primers used for PCR, cloning, and mutagenesis are listed in Supplementary Table 1. *tarP* (UniProt A0A0H3JNB0, NCBI Gene ID 1260584) was amplified using genomic DNA of *S. aureus* N315 and inserted in *Escherichia coli*/*S. aureus* shuttle vector pRB474[38] at the BamHI and SacI sites, to transform *S. aureus*, or into pQE80L at BamH1 and HindIII sites, to transform *E. coli* BL21(DE3). A thrombin cleavage site was inserted between the His-tag and mature protein in pQE80L. Single mutations of TarP were introduced by PCR-based site-directed mutagenesis[39]. The obtained amplicons were confirmed by sequencing. For the construction of marker-less *S. aureus* deletion mutants of *tarS* or *tarP*, the pIMAY shuttle vector was used[40]. The IgG-binding surface protein A gene (*spa*) was deleted using the pKORI shuttle vector[41]. Protein A deletion had

no impact on phage siphophage or podophage susceptibility, indicating that it did not alter WTA amount or structure.

Protein expression, purification, and activity assay. *E. coli* BL21(DE3) were grown in LB medium at 30 °C. Expression of *tarP* was induced with 1 mM IPTG at 22 °C at an OD_{600} of 0.6. After 15 h, cells were harvested, washed with wash buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA), and lysed by sonication with lysis buffer (70 mM NaH_2PO_4, pH 8.0, 1 M NaCl, 20% glycerol, 10 U/ml of benzonase nuclease). After centrifugation (15,000*g*), the supernatant was filtered with a 0.45 μm filter, loaded onto a His Trap FF column (GE Healthcare, 5 ml), and washed with buffer A (50 mM NaH_2PO_4, pH 8.0, 1 M NaCl, 20% glycerol) supplemented with 45 mM imidazole and buffer B (buffer A with 90 mM imidazole). Finally, the protein was eluted with buffer C (buffer A with 500 mM imidazole), and the fractions were pooled, and further purified by size-exclusion chromatography on a Superdex 200 10/30 column equilibrated with buffer D (20 mM MOPS, pH 7.6, 400 mM LiCl, 10 mM $MgCl_2$, 5 mM β-mercaptoethanol, 5% glycerol). The peak fractions were pooled and concentrated to 1.4 mg/ml for crystallization. For selenomethionyl-form TarP production, bacteria were grown in a selenomethionine-containing medium (Molecular Dimension) and auto-induction was carried out. The protein was purified as described above. The activity of wild-type and mutated TarP, as well as donor substrate specificity of TarP were determined with the ADP Quest Assay kit (DiscoverRx, Extended Data Tables 2, 3). The reaction volume was 20 μl with 1 mM UDP-GlcNAc, 1.5 mM purified WTA from RN4220 Δ*tarM/S*. The reaction was started with protein and incubated at room temperature for 1 h. Released UDP, coupled into a fluorescence signal, was detected in a 384-well black assay plate with 530 nm excitation and 590 nm emission wavelengths using TECAN Infinite M200.

Crystallization and data collection. Crystals were obtained by vapour diffusion at 20 °C. 1 μl protein solution was mixed with 1 μl reservoir solution containing 25% PEG 3350, 250 mM $MgCl_2$, and 0.1 M sodium citrate, pH 5.7. The selenothionyl-form protein was crystallized under the same conditions. For crystals of TarP with UDP-GlcNAc, 27 mM UDP-GlcNAc was introduced in the reservoir solution containing 250 mM $MgCl_2$ or 230 mM $MnCl_2$. Crystals of TarP with Mg^{2+} were used for soaking of synthetic 3RboP (60 mM), 6RboP-$(CH_2)_6NH_2$ (41 mM), or UDP-GlcNAc (20 mM) combined with 3RboP (52 mM) for 5 min. For data collection the crystals were cryo-protected with 20% glycerol in reservoir solution and flash-frozen in liquid nitrogen. Diffraction data were collected at beamline X06DA of Swiss Light Source in Villigen, Switzerland, or at beamline BL14.1 at BESSY-II, Helmholtz Zentrum Berlin.

Phasing, model building, and refinement. For phase determination, two data sets from a selenomethionine-containing TarP crystal were collected at wavelengths of 0.97941 Å (peak) and 0.97952 Å (inflection). The structure was solved by multiwavelength anomalous dispersion (MAD) at 2.60 Å resolution. All data were reduced using XDS/XSCALE software packages[42]. Initial phases were derived from the substructure of 26 selenium atom sites per asymmetric unit with the program suite SHELX C/D/E[43]. The heavy atom parameters were further refined and the initial phases were improved by SHARP/autoSHARP[44]. The initial model was generated with PHENIX[45] and the final model was achieved by cycles of iterative model modification using COOT[46], and restrained refinement with REFMAC. TLS was used in the later stages[47,48]. The four binary and one ternary complex structures were solved by molecular replacement with PHASER[49] and the unliganded TarP structure was used as a search model. UDP-GlcNAc, 3RboP, Mg^{2+}, or Mn^{2+} were removed from the models to calculate the simulated annealing ($mF_o − DF_c$) omit maps using PHENIX. The anomalous difference map of Mn^{2+} at 1.89259 Å was generated by FFT within CCP4, from which two Mn^{2+} in the active site and one Mn^{2+} at the trimer interface were identified. The coordinate and parameter files for 3RboP and 6RboP-$(CH_2)_6NH_2$ were calculated using the PRODRG server[50]. The structure figures were generated by PyMOL[51] and the models were evaluated using MolProbity[52]. Statistics for the data collection, phasing, and refinement are reported in Extended Data Tables 4 and 5.

Synthesis of ribitol phosphate oligomers. *Synthesis of 3RboP.* Target compound **1**, D-ribitol-5-phosphate trimer (3RboP), was prepared by the phosphoramidite method[53,54] (Supplementary Fig. 2). In brief, the primary alcohol of commercially available compound **2** was converted into levulinoyl ester by using levulinic acid and *N,N'*-dicyclohexylcarbodiimide (DCC), and the allyl group of **3** was removed with tetrakis(triphenylphosphine)palladium to produce compound **4**. The primary alcohol of **4** reacted with phosphine derivative **5** in the presence of diisopropylammonium tetrazolide[55] to generate phosphoramidite **6**. At the same time, compound **4** was coupled with dibenzyl *N,N*-diisopropylphosphoramidite **7**, which was catalysed by 1*H*-tetrazole, and the product was further oxidized by *tert*-butyl hydroperoxide, yielding protected D-ribitol-5-phosphate **8**. Cleavage of the levulinoyl ester of **8** with hydrazine hydrate resulted in benzyl protected D-ribitol-5-phosphate **9**, which was further coupled with phosphoramidite **6** and oxidized with *tert*-butyl hydroperoxide to yield protected dimers of D-ribitol-5-phosphate **10**. After removal of the levulinoyl group, the dimer **11** was coupled with phosphoramidite **6** using the same conditions as above to obtain a protected trimer of D-ribitol-5-phosphate

12. Subsequent removal of the levulinoyl group and hydrogenolysis of **13** to remove all benzyl groups yielded 3RboP **1**. All chemicals and experimental procedures as well as characterization of products can be found in the Supplementary Methods. *Synthesis of 6RboP-(CH₂)₆NH₂.* Aminohexyl D-ribitol-5-phosphate hexamer (6RboP-(CH₂)₆NH₂) was synthesized using a new method (Supplementary Fig. 3). All chemicals (Acros, Biosolve, Sigma-Aldrich and TCI) for the synthesis were used as received and all reactions were performed under a protective argon atmosphere at room temperature, unless otherwise stated. Procedures for phosphoramidite coupling, oxidation, detritylation, global deprotection, TLC analysis and characterization of these compounds can be found in Supplementary Methods.

Human samples. Venous blood samples were obtained from male and female healthy volunteers (20–50 years) with protocols approved by the Institutional Review Board for Human Subjects at the University of Tübingen (014/2014BO2 und 549/2018BO2). Informed written consent was obtained from all volunteers. Blood samples were used for purification of either serum IgGs or neutrophils as described below.

IgG from human plasma. IgG was purified from plasma of human donors using the NAb Protein G Spin Kit (ThermoFisher), purity was checked by SDS PAGE, and protein concentration was determined using Bradford assay. Anti-WTA-IgG was prepared as described[9]. To analyse the IgG-binding capacity of *S. aureus* cells, exponentially growing bacterial cultures were adjusted to an OD_{600} of 0.5, diluted 1:10 in PBS, and 100 μl of diluted bacteria was mixed with 100 μl of IgG diluted in PBS with 1% BSA. The concentration of IgG was 250 ng/ml for IgG enriched for WTA binding, 10 μg/ml for IgG from pooled human serum (Athens R&T 16-16-090707, Abcam ab98981), or 5 μg/ml for single-donor IgG preparations. A control without IgG was included in all experiments for all mutants. Samples were incubated at 4 °C for 1 h, centrifuged, washed 2–3 times with PBS, and further incubated with 100 μl FITC-labelled anti-human IgG (Thermo Scientific, 62-8411, 1:100 in PBS with 1% BSA, 62-8411) at 4 °C for 1 h. Bacteria were centrifuged, washed 2–3 times with PBS, and fixed with 2% paraformaldehyde (PFA). Surface-bound IgG was quantified by flow cytometry using a BD FACSCalibur. For all flow cytometry experiments a mutant panel lacking *spa*, the gene for the IgG-binding protein A, was used. The subsequent gating strategy is exemplified in Extended Data Fig. 5a.

IgG-mediated phagocytosis. Stationary-phase *S. aureus* cells were washed once with PBS and labelled by incubation in PBS containing 10 μM carboxyfluorescein succinimidyl ester (CFSE; OD_{600} of 1.7) at 37 °C for 1 h. The bacteria were washed three times and resuspended in PBS. CFU were determined by plating on TSB plates and bacteria were heat-inactivated at 70 °C for 20 min. CFSE-labelled *S. aureus* (1×10^7 cells/ml) in PBS with 0.5% BSA were opsonized with anti-WTA-IgG (0.15 or 0.3 ng/μl) at 4 °C for 40 min. Neutrophils from human donors, isolated via Ficoll-Histopaque density gradient centrifugation[56], were diluted to a concentration of 2.5×10^6/ml in neutrophil medium (10% HSA, 2 mM L-glutamine, 2 mM sodium pyruvate, 10 mM HEPES). 200 μl neutrophil suspension was incubated with 25 μl opsonized bacteria (final MOI 0.5) in a 96-well plate at 37 °C for 30 min, centrifuged (350*g*, 10 min), washed once with 200 μl PBS, and fixed with 2% PFA at room temperature for 15 min. Cells were washed twice with PBS and analysed by flow cytometry, whereby surface-bound and ingested bacteria were measured without discrimination. An example of the neutrophil gating strategy can be found in Extended Data Fig. 5b.

Mice. Six-week-old sex-matched wild-type C57BL/6J mice, purchased from ORIENT BIO (Charles River Breeding Laboratories in Korea), were kept in micro-isolator cages in a pathogen-free animal facility. The conducted experiments were performed according to guidelines and approval (PNU-2017-1503) by the Pusan National University-Institutional Animal Care and Use Committee (PNU-IACUC). Sample size was chosen to obtain significant outcomes (alpha error ≤ 5%), based on results from previous experiments[21]. Experiments were performed in a non-blinded, non-randomized fashion.

Mouse vaccination and infection. 30 μg of purified WTA from *S. aureus* N315 wild-type or isogenic Δ*tarP*, or Δ*tarS* mutants was dissolved in 15 μl PBS and mixed with the same volume of aluminium hydroxide gel adjuvant (Alhydrogel 1.3%, 6.5 mg/ml, Brennatag). The mixtures were incubated at 37 °C with agitation for 1 h and injected three times at one-week intervals via mouse footpads. Seven days after the third injection, blood was obtained from the retro-orbital sinus and centrifuged (9,000*g*) at 4 °C for 10 min. The supernatants were aliquoted (50 μl) and stored at –80 °C for ELISA quantification of WTA-binding IgG as described[57]. Sera were diluted 1:100 and tested by ELISA on 96-well plates coated with 2.5 μg/ml sonicated WTA preparations (WTA from N315, Δ*tarS* or Δ*tarP*, respectively).

To prepare an inoculum for infection, N315 wild-type bacteria were grown in TBS at 37 °C with agitation (180 r.p.m.) until they reached an OD_{600} of 1.0. After centrifugation (3,500*g*) at 4 °C for 10 min, bacteria adjusted to 5×10^7 CFU in 50 μl PBS containing 0.01% BSA were intravenously injected ($n = 5$ per group). Injected bacterial numbers were verified by plating serial dilutions of the inoculum onto TSA plates. To determine residual bacterial dissemination to kidneys, challenged

mice were euthanized, and organs were extracted aseptically and homogenized in 1 ml of saline using a Polytron homogenizer (PT3100). The homogenates were serially diluted and plated on TSA to determine CFU counts. CFU were calculated per 1 ml of kidney.

Statistical analyses. Statistical analysis was performed by using GraphPad Prism (GraphPad Software, Inc.). Statistically significant differences were calculated by appropriate statistical methods as indicated. *P* values of ≤ 0.05 were considered significant.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All major data generated or analysed in this study are included in the article or its supplementary information files. The coordinates and structure factors were deposited in the Protein Data Bank under accession numbers 6H1J, 6H21, 6H2N, 6H4F, 6H4M and 6HNQ. Source data for experiments with animals (Fig. 4c, d) are provided. Additionally, a gel image of Extended Data Fig. 1f is supplied as Supplementary Fig. 1. All other data relating to this study are available from the corresponding authors on reasonable request.

28. Schulte, B., Bierbaum, G., Pohl, K., Goerke, C. & Wolz, C. Diversification of clonal complex 5 methicillin-resistant *Staphylococcus aureus* strains (Rhine-Hesse clone) within Germany. *J. Clin. Microbiol.* **51**, 212–216 (2013).
29. Larsen, J. et al. Meticillin-resistant *Staphylococcus aureus* CC398 is an increasing cause of disease in people with no livestock contact in Denmark, 1999 to 2011. *Euro Surveill.* **20**, 30021 (2015).
30. Sieber, R. N. et al. Drivers and dynamics of methicillin-resistant livestock-associated *Staphylococcus aureus* CC398 in pigs and humans in Denmark. *mBio* **9**, e02142-18 (2018).
31. European Committee for Antimicrobial Susceptibility Testing (EUCAST) of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID). Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution. *Clin. Microbiol. Infect.* **9**, ix–xv (2003).
32. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
33. Winstel, V., Sanchez-Carballo, P., Holst, O., Xia, G. & Peschel, A. Biosynthesis of the unique wall teichoic acid of *Staphylococcus aureus* lineage ST395. *MBio* **5**, e00869 (2014).
34. Tormo, M. A. et al. *Staphylococcus aureus* pathogenicity island DNA is packaged in particles composed of phage proteins. *J. Bacteriol.* **190**, 2434–2440 (2008).
35. Chen, P. S., Toribara, T. Y. & Warner, H. Microdetermination of phosphorus. *Anal. Chem.* **28**, 1756–1758 (1956).
36. Smith, R. L. & Gilkerson, E. Quantitation of glycosaminoglycan hexosamine using 3-methyl-2-benzothiazolone hydrazone hydrochloride. *Anal. Biochem.* **98**, 478–480 (1979).
37. Xia, G. et al. Glycosylation of wall teichoic acid in *Staphylococcus aureus* by TarM. *J. Biol. Chem.* **285**, 13405–13415 (2010).
38. Brückner, R. A series of shuttle vectors for *Bacillus subtilis* and *Escherichia coli*. *Gene* **122**, 187–192 (1992).
39. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).
40. Monk, I. R., Shah, I. M., Xu, M., Tan, M. W. & Foster, T. J. Transforming the untransformable: application of direct transformation to manipulate genetically *Staphylococcus aureus* and *Staphylococcus epidermidis*. *MBio* **3**, e00277-11 (2012).
41. Bae, T. & Schneewind, O. Allelic replacement in *Staphylococcus aureus* with inducible counter-selection. *Plasmid* **55**, 58–63 (2006).
42. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
43. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
44. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
45. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
47. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
48. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
49. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
50. Schüttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363 (2004).
51. Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* (2015).
52. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

53. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedr. Lett.* **22**, 1859–1862 (1981).
54. Elie, C. J. J. et al. Synthesis of fragments of the capsular polysaccharide of *Haemophilus influenzae* type b: Part IIII-3. A solid-phase synthesis of a spacer-containing ribosylribitol phosphate hexamer. *Recl. Trav. Chim. Pays Bas* **108**, 219–223 (1989).
55. Dreef, C. E., Elie, C. J. J., Hoogerhout, P., van der Marel, G. A. & van Boom, J. H. Synthesis of 1-O-(1,2-di-O-palmitoyl-*sn*-glycero-3-phospho)-d-myo-inositol 4,5-bisphosphate: an analogue of naturally occurring (ptd)Ins(4,5)P$_2$. *Tetrahedr. Lett.* **29**, 6513–6515 (1988).
56. Dürr, M. C. et al. Neutrophil chemotaxis by pathogen-associated molecular patterns—formylated peptides are crucial but not the sole neutrophil attractants produced by *Staphylococcus aureus*. *Cell. Microbiol.* **8**, 207–217 (2006).
57. Caulfield, M. J. et al. Small molecule mimetics of an HIV-1 gp41 fusion intermediate as vaccine leads. *J. Biol. Chem.* **285**, 40604–40611 (2010).
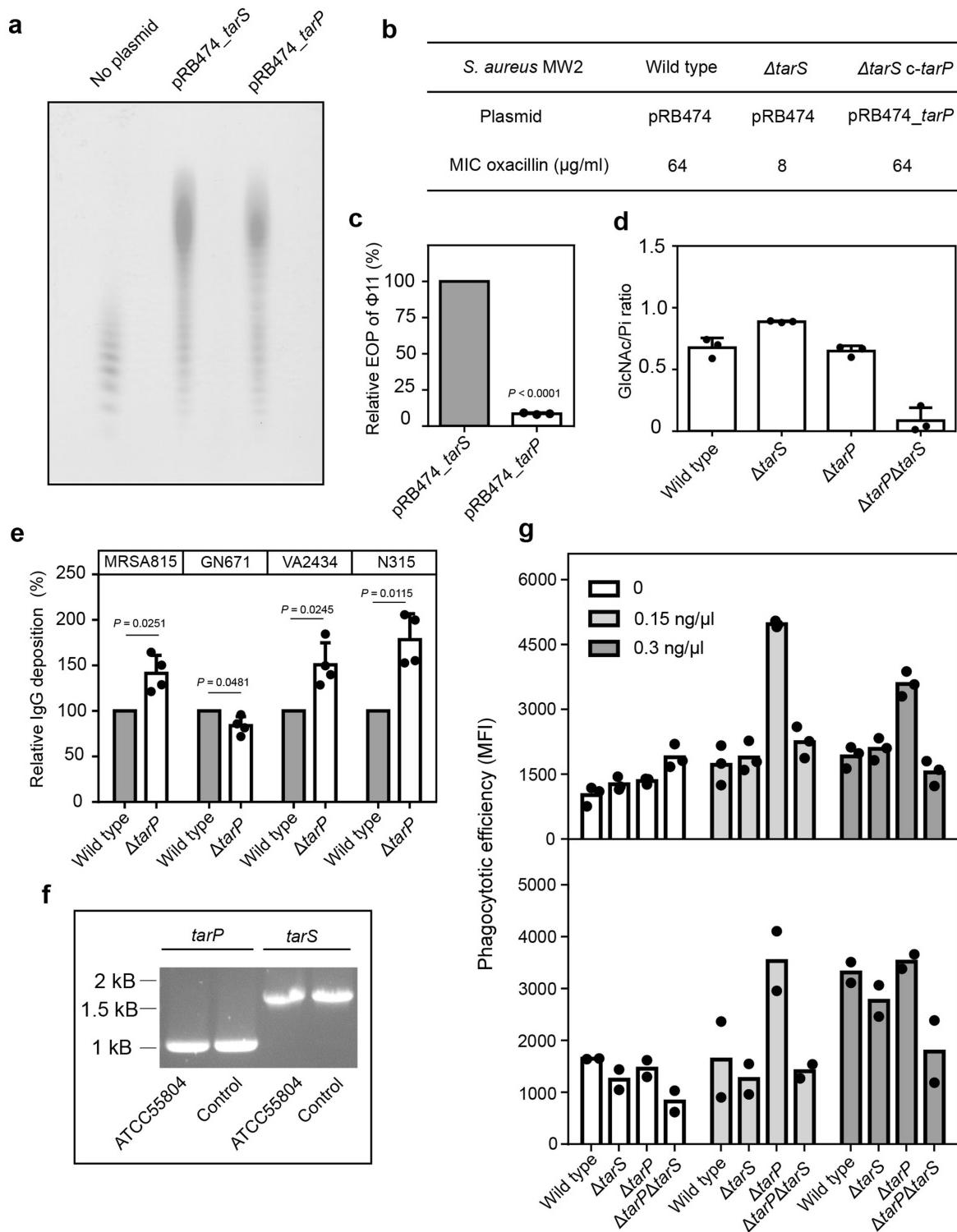
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Characterization of TarP, deposition of human IgGs, and presence of *tarP* in the producer of antigen 336. a, Analysis of WTA by PAGE. WTA from RN4220 Δ*tarM/S* expressing either *tarP* or *tarS* was compared with non-glycosylated WTA. Data shown are representative of two experiments. **b**, MIC values of oxacillin against MW2 wild type, *tarS* mutant, and *tarP*-complemented *tarS* mutant. Data are medians of ten independent experiments. **c**, Efficiency of plating (EOP) of phage Φ11 against *tarS* or *tarP*-expressing RN4420 Δ*tarM/S*. Values of *tarP* relative to *tarS* expression are given as mean ± s.d. ($n = 3$). Statistical significance was calculated by paired Student's *t*-test (two-sided) with significant *P* values ($P \leq 0.05$) indicated. **d**, The level of WTA glycosylation catalysed by TarP or TarS was determined by analysing the GlcNAc and phosphate content of WTA isolated from a N315 strain panel. Depicted is the ratio of GlcNAc and phosphate as mean with s.d. of three technical replicates. The values are in good agreement with NMR data (Supplementary Table 3). **e**, Relative deposition of IgG from intravenous immunoglobulins enriched for WTA binding on different CC5 wild-type and *tarP* mutant cells. Values are given as mean percentage ± s.d. of four independent experiments. Statistical significance was calculated by paired Student's *t*-test (two-sided). *P* values ≤ 0.05 were considered significant and are indicated. **f**, Presence of *tarP* and *tarS* in *S. aureus* ATCC55804, expressing antigen 336, described as 3-O-GlcNAc-WTA[25]. Shown is a representative of two independent replicates. **g**, TarP reduces neutrophil phagocytosis of N315 strains lacking protein A, opsonized with indicated concentrations of IgG enriched for WTA binding. Values are depicted as mean fluorescence intensity (MFI). Shown are two independent experiments with neutrophils from different donors. These data supplement data presented in Fig. 4b: upper panel, mean of three technical replicates of an independent experiment, lower panel, mean of two technical replicates.

Extended Data Fig. 2 | NMR analysis of WTA from N315 mutant panel. All depicted experiments were repeated twice. *y*-axes and *x*-axes show ^{13}C and ^{1}H chemical shifts, respectively. **a–d**, NMR spectra of non-glycosylated WTA (Δ*tarS*Δ*tarP* mutant). **a**, HSQC expansion of the region containing the ribitol and glycerol protons shifted by acylation; **b, c**, HSQC-TOCSY-20 and HSQC-TOCSY-80 spectra, respectively. **d**, HSQC area of the non-acylated ribitol and glycerol proton. **e–h**, NMR spectra of TarS-WTA (Δ*tarP* mutant). **e**, HSQC expansion of the region containing the ribitol and glycerol protons shifted by acylation. **f, g**, HSQC-TOCSY-20 and HSQC-TOCSY-80, respectively. **h**, HSQC area of the non-acylated ribitol and glycerol proton. **i–o**, NMR spectra of TarP-WTA (Δ*tarS* mutant). **i**, HSQC expansion of the region containing the

ribitol and glycerol protons shifted by acylation. **j, k**, HSQC-TOCSY-20 and HSQC-TOCSY-80 spectra, respectively. **l**, HSQC area of the non-acylated ribitol and glycerol protons. **m**, Expansion of **l** with HSQC (black/grey) overlapped with HSQC-TOCSY-20 (cyan). **n**, Overlap of HSQC-TOCSY-20 (cyan) and HSQC-TOCSY-80 (black). **o**, HSQC (black) and HMBC (grey) detailing the GlcNAc signals. **p**, NOESY expansion detailing the correlations of the β-GlcNAc anomeric protons: GlcNAc 'b*' differs from unit 'b', which has the same anomeric proton chemical shift, but is linked to a different ribitol unit. All densities are labelled with the letters used in Supplementary Table 2. The density marked with an asterisk in **m** is consistent with ribitol glycosylated at O-4.
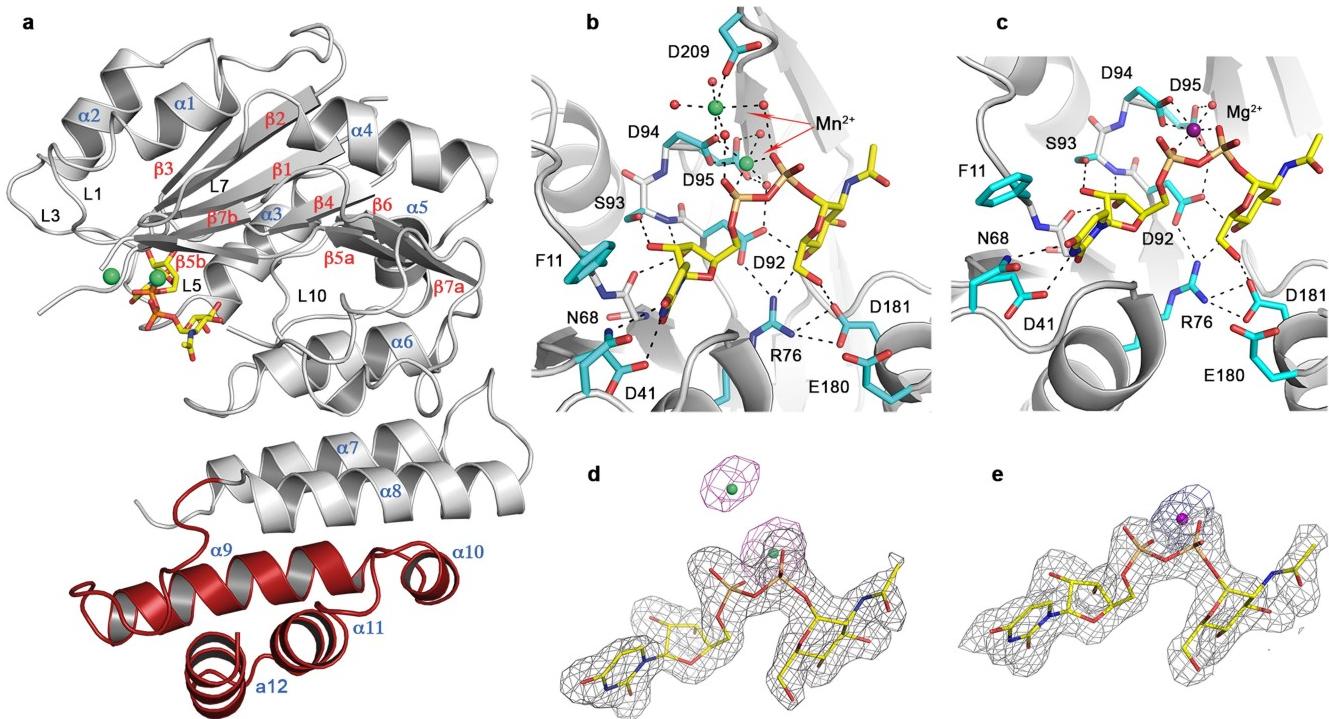
Extended Data Fig. 3 | Secondary structure of a TarP monomer and interactions with UDP-GlcNAc. a, Cartoon representation of a TarP monomer bound to UDP-GlcNAc (yellow) and Mn^{2+} (lime green). The CTD is coloured red. **b**, Interactions of TarP with UDP-GlcNAc and Mn^{2+}, coloured as in **a**. Hydrogen bonds and salt bridges are shown as black dashed lines. **c**, Interactions of TarP with UDP-GlcNAc (yellow) and Mg^{2+} (magenta). **d**, Simulated-annealing ($mF_o - DF_c$) omit map of

UDP-GlcNAc (grey mesh, contoured at 2.0σ) and Mn^{2+} (magenta mesh, at 3.0σ) in the TarP–UDP-GlcNAc–Mn^{2+} complex structure. UDP-GlcNAc and Mn^{2+} are coloured as in **a**. **e**, Simulated-annealing ($mF_o - DF_c$) omit map of UDP-GlcNAc (grey mesh, at 2.0σ) and Mg^{2+} (blue mesh, at 2.0σ) in the TarP–UDP-GlcNAc–Mg^{2+} complex structure. UDP-GlcNAc and Mg^{2+} are coloured as in **c**.
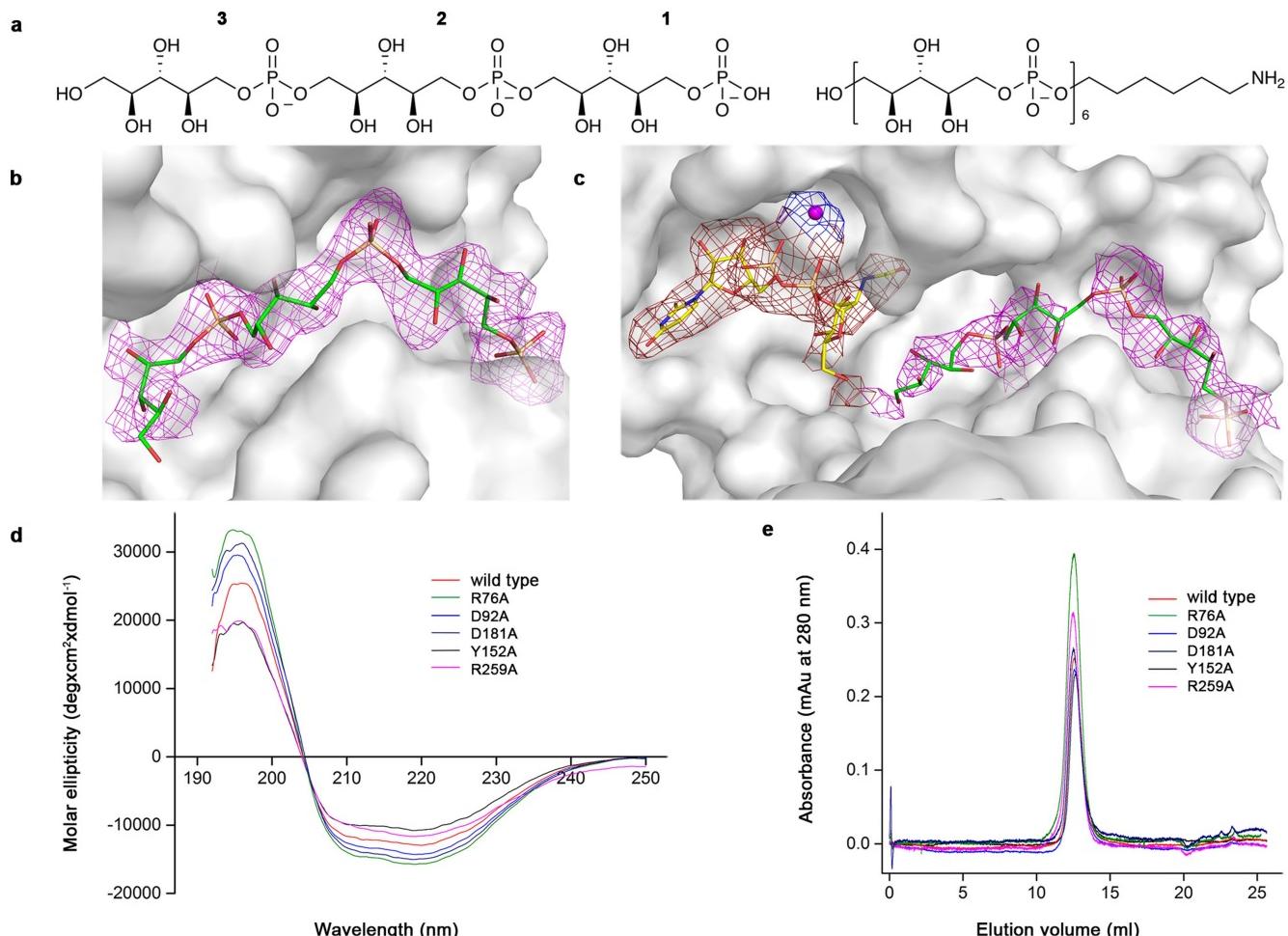
Extended Data Fig. 4 | Simulated-annealing ($mF_o − DF_c$) omit maps of 3RboP and UDP-GlcNAc, and characterization of TarP mutant proteins. a, Chemical structures of synthetic 3RboP and 6RboP-$(CH_2)_6NH_2$. The unit numbers are indicated. **b**, Simulated-annealing ($mF_o − DF_c$) omit map of 3RboP (lime green) in the binary structure (magenta mesh, contoured at 2.0σ). **c**, Simulated-annealing ($mF_o − DF_c$) omit map of UDP-GlcNAc (yellow), Mg^{2+} (magenta) and 3RboP (lime green) in the ternary complex structure (red mesh, at 1.8σ, blue mesh, at 2.0σ or magenta mesh, at 1.5σ). **d**, Circular dichroism spectra of wild-type and mutant TarP proteins (for wild type, R76A and D181A, $n = 3$; for D92A, Y152A and R259A, $n = 2$). **e**, Size-exclusion chromatography elution profiles of wild-type and mutant TarP proteins (for wild type, $n = 8$; for R76A, D181A and R259A, $n = 3$; for D92A and Y152A, $n = 2$, all with similar results). Mutant proteins D94A, E180A, D209A, K255A, R262A, and H263A showed similar circular dichroism spectra and size-exclusion chromatography elution profiles (data not shown).

a

b
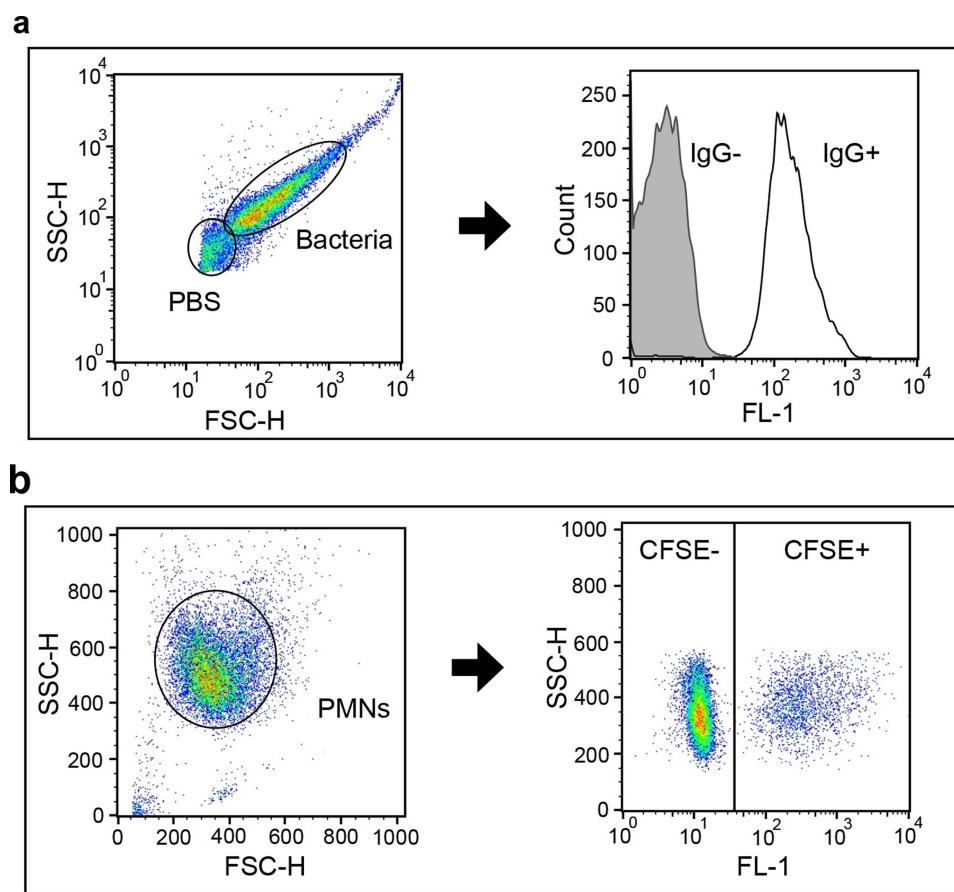
Extended Data Fig. 5 | Gating strategy for flow cytometry experiments.
a, Gating strategy for IgG deposition experiments. To distinguish bacteria from background signals, pure PBS was measured. Left, bacterial gating occurred at the FSC/SCC density plot omitting PBS-derived signals. Bacterial aggregates of high SSC and FSC values were excluded from the gated population as well. Right, the mean fluorescence of the bacterial population (black) was determined and compared with non-IgG-treated bacteria (grey) to control for nonspecific binding of the secondary FITC-labelled antibody. Subsequently, mean fluorescence values of individual mutants were compared relatively to the corresponding wild-type strain.

b, Gating strategy for phagocytosis experiments. Neutrophils were separated by Histopaque/Ficoll gradient and subsequent gating of neutrophils occurred at the FSC/SCC density plot upon size and complexity (left). Histopaque/Ficoll gradient isolations showed a neutrophil purity of more than 80%. Using the CFSE-fluorescence channel, the gated population was subdivided into fluorescence-positive and -negative cells (right). Successful phagocytosis was indicated by uptake of CFSE-labelled bacteria. The phagocytic efficiency was expressed as product of the mean fluorescence of the fluorescence-positive population and their relative abundance (mean fluorescence intensity, MFI).

Extended Data Table 1 | *tarP* presence and podophage susceptibility of CC5 strains, comprising sequence type (ST) 5 and 225, and CC398 isolates

Collection	Rhine-Hesse collection		Danish LA-MRSA collection		MRSA collection Tübingen	
Clonal complex	5 (ST5 + ST225)		398		5 (ST5 + ST225)	
tarP status	Negative	Positive	Negative	Positive	Negative	Positive
n	21	39	18	12	11	53
Phage susceptibility	Susceptible	Resistant	Susceptible	Resistant	Susceptible	Resistant
φ44	21	39	18	12	ND	ND
φ66	21	39	18	12	ND	ND
φP68	21	39	18	12	ND	ND

tarP presence in three different *S. aureus* collections was determined by PCR using primer pair TarP_Ty_Fw/Rv. Phage susceptibility to podophages Φ44, Φ66, and ΦP68 was determined by soft-agar overlay. Plaque formation indicated susceptibility, absence of visible plaque formation indicated resistance. ND, not determined.

Extended Data Table 2 | Enzymatic activities of mutated TarP proteins

Function	TarP variant	Activity of wild type in %
Trimer interface	I322E	128
UDP-GlcNAc binding	R76A	1
	D92A	2
	D94A	14
	D209A	105
Catalytic base	E180A	15
	D181A	1
3RboP binding	Y152A	44
	K255A	99
	R259A	3
	R262A	97
	H263A	81

Extended Data Table 3 | Donor substrate specificity of TarP

Sugar nucleotide	Enzymatic activity (nmol/mg*min)
UDP-GlcNAc	2.20
UDP-Glc	0.01
UDP-GalNAc	0.03
UDP-Gal	0.01

Extended Data Table 4 | Crystallographic data statistics for TarP and TarP–UDP-GlcNAc–Mg^{2+}

	TarP native	TarP-SeMet	TarP-SeMet	TarP-UDP-GlcNAc-Mg^{2+}
		Peak	*Inflection*	
Data collection				
Space group	P2$_1$	P2$_1$	P2$_1$	P2$_1$
Cell dimensions				
a, b, c (Å)	43.37, 95.25, 125.47	44.06, 95.33, 130.72	43.99, 95.22, 130.52	43.85, 95.27, 130.22
α, β, γ (°)	90.00, 96.57, 90.00	90.00, 93.41, 90.00	90.00, 93.34, 90.00	90.00, 93.49, 90.00
Wavelength (Å)	1.00004	0.97941	0.97952	0.91841
Resolution (Å)	44.5-1.86 (1.91-1.86)	47.7-2.29 (2.35-2.29)	47.7-2.30 (2.35-2.30)	47.6-1.95 (2.00-1.95)
R_{sym} or R_{merge} (%)	8.4 (87.7)	11.5 (103.8)	9.7 (62.2)	12.6 (110.1)
I / $\sigma(I)$*	9.4 (1.4)	13.8 (1.8)	15.8 (2.9)	9.2 (1.3)
CC$_{1/2}$ (%)	99.7 (50.0)	99.8 (64.0)	99.8 (81.9)	99.6 (50.6)
Completeness (%)	98.5 (97.5)	99.0 (88.4)	99.2 (90.9)	99.9 (99.7)
Redundancy	2.9 (2.7)	7.0 (6.5)	6.6 (6.0)	5.0 (5.0)
Phasing				
Rcullis (ano)		0.76		
Phasing power		1.24		
HA sites / ASU		26		
FOM$_{acentric}$		0.41		
Refinement				
Resolution (Å)	44.5 - 1.86			47.6 - 1.95
No. reflections	241855 (16740)			386853 (28878)
R_{work} / R_{free} (%)	17.1/21.8			17.7/22.4
No. atoms				
Protein	7538			7479
Substrates	0			117
Ions	13			29
Other molecules	0			24
Water	697			804
Average B-factors (Å^2)				
Protein	31.7			35.5
Substrates				43.9
Ions	40.1			44.6
Other molecules				39.2
Water	41.6			41.0
R.m.s deviations**				
Bond lengths (Å)	0.010			0.008
Bond angles (°)	1.310			1.254
Ramachandran plot				
Favored (%)	97			97
Allowed (%)	3			3
Outliers (%)	0			0

Values in parentheses are for highest-resolution shell. Two data sets for TarP-SeMet were collected from the same single crystal.
*I is the mean intensity, $\sigma(I)$ is the standard deviation of reflection intensity I.
**r.m.s.d., root-mean-square deviation of bond length or bond angle.

Extended Data Table 5 | Crystallographic data statistics for TarP–UDP-GlcNAc–Mn^{2+}, TarP–3RboP, TarP–6RboP-(CH$_2$)$_6$NH$_2$ and TarP–UDP-GlcNAc–3RboP

	TarP-UDP-GlcNAc-Mn^{2+}	TarP-3RboP	TarP-6RboP-(CH$_2$)$_6$NH$_2$	TarP-UDP-GlcNAc-3RboP
Data collection				
Space group	P2$_1$	P2$_1$	P2$_1$	P2$_1$
Cell dimensions				
a, b, c (Å)	43.86, 95.36, 130.55	95.61, 217.27, 123.99	95.41, 211.25, 122.68	95.17, 210.75, 123.20
α, β, γ (°)	90.00, 93.51, 90.00	90.00, 91.38, 90.00	90.00, 91.61, 90.00	90.00, 91.92, 90.00
Wavelength (Å)	0.91840	1.00000	1.00002	1.00002
Resolution (Å)	47.7-1.80 (1.85-1.80)	49.8-2.16 (2.22-2.18)	48.5-2.40 (2.46-2.40)	48.4-2.73 (2.80-2.73)
R_{sym} or R_{merge} (%)	5.6 (101.0)	13.7 (140.9)	15.6 (141.2)	25.4 (161.1)
$I / \sigma(I)$*	12.0 (1.3)	11.9 (1.5)	10.8 (1.5)	8.4 (1.4)
CC$_{1/2}$ (%)	99.9 (51.1)	99.8 (54.0)	99.6 (50.7)	99.0 (52.3)
Completeness (%)	99.8 (99.5)	100.0 (100.0)	99.9 (100.0)	99.9 (99.8)
Redundancy	3.6 (3.3)	7.0 (6.6)	6.2 (6.4)	7.1 (7.4)
Refinement				
Resolution (Å)	47.7 - 1.80	49.8 - 2.18	48.5 - 2.40	48.4 - 2.73
No. reflections	355981 (24195)	1833608 (128618)	1172903 (89756)	911354 (69899)
R_{work} / R_{free} (%)	17.6/21.3	17.1/20.7	19.6/22.7	19.2/23.5
No. atoms				
Protein	7,543	29,987	29,709	29,439
Substrates	117	480	480	948
Ions	19	32	16	35
Other molecules	12	18		
Water	739	2,694	1,555	1,383
Average B-factors (Å^2)				
Protein	37.6	46.1	51.2	53.0
Substrates	38.4	57.8	75.0	84.3
Ions	47.4	52.7	54.0	50.6
Other molecules	46.6	49.7		
Water	43.7	49.4	48.6	41.4
R.m.s deviations**				
Bond lengths (Å)	0.010	0.009	0.008	0.010
Bond angles (°)	1.331	1.288	1.214	1.302
Ramachandran plot				
Favored (%)	98.0	97.0	96.8	96.4
Allowed (%)	2.0	3.0	3.2	3.6
Outliers (%)	0	0	0	0

Values in parentheses are for highest-resolution shell.
*I is the mean intensity, $\sigma(I)$ is the standard deviation of reflection intensity I.
**r.m.s.d., root-mean-square deviation of bond length or bond angle.

LETTER

Cryptic connections illuminate pathogen transmission within community networks

Joseph R. Hoyt[1,2]*, Kate E. Langwig[2], J. Paul White[3], Heather M. Kaarakka[3], Jennifer A. Redell[3], Allen Kurta[4], John E. DePue[5], William H. Scullon[6], Katy L. Parise[7,8], Jeffrey T. Foster[7,8], Winifred F. Frick[1,9] & A. Marm Kilpatrick[1]

Understanding host interactions that lead to pathogen transmission is fundamental to the prediction and control of epidemics[1-5]. Although the majority of transmissions often occurs within social groups[6-9], the contribution of connections that bridge groups and species to pathogen dynamics is poorly understood[10-12]. These cryptic connections—which are often indirect or infrequent—provide transmission routes between otherwise disconnected individuals and may have a key role in large-scale outbreaks that span multiple populations or species. Here we quantify the importance of cryptic connections in disease dynamics by simultaneously characterizing social networks and tracing transmission dynamics of surrogate-pathogen epidemics through eight communities of bats. We then compared these data to the invasion of the fungal pathogen that causes white-nose syndrome, a recently emerged disease that is devastating North American bat populations[13-15]. We found that cryptic connections increased links between individuals and between species by an order of magnitude. Individuals were connected, on average, to less than two per cent of the population through direct contact and to only six per cent through shared groups. However, tracing surrogate-pathogen dynamics showed that each individual was connected to nearly fifteen per cent of the population, and revealed widespread transmission between solitarily roosting individuals as well as extensive contacts among species. Connections estimated from surrogate-pathogen epidemics, which include cryptic connections, explained three times as much variation in the transmission of the fungus that causes white-nose syndrome as did connections based on shared groups. These findings show how cryptic connections facilitate the community-wide spread of pathogens and can lead to explosive epidemics.

Pathogens have repeatedly spilled over from wildlife to humans and caused epidemics at local to global scales[16]. Social behaviour often constrains the transmission of pathogens, with network structure increasing transmission within social groups[1,7,12] but limiting the spread to solitary individuals, other social groups, and between species[8,10,17]. Contacts that bridge species and social groups are usually infrequent or indirect—for example, via the environment—but they can lead to unexplained cases and outbreaks[18] including the 2014 Ebola epidemic in West Africa[19], Nipah virus transmission in Bangladesh[18], transmission of *Mycobacterium bovis* in badgers[20] and *Mycoplasma* in tortoises[21]. For white-nose syndrome, an emerging infectious disease of bats, impacts are exceedingly high in some species that roost solitarily[14], which raises the important question of how solitary individuals acquire infections.

Cryptic connections are important because they can determine whether transmission is constrained to single-species, local outbreaks or spans multiple populations and species. However, the extent of cryptic connections among social groups and their influence on pathogen dynamics is poorly known, in part, because of the difficulty in measuring these types of contacts. Prospective studies of contact rates

often take place over short time periods, are usually designed to capture direct connections among individuals and can miss infrequent or indirect connections[3,22-24]. Retrospective methods such as contact tracing[23,24] and model estimation of contact rates[17] sometimes uncover cryptic connections, but determining the general importance of these connections in transmission is often limited by epidemic-specific details[25].

Here we examine how cryptic connections between social groups and species influence pathogen transmission in replicated epidemics in multiple populations of hibernating bats. As with many other species[20,26], hibernating bats spend the majority of their time in relatively small groups[14]. However, individual bats arouse from hibernation for short periods (1–3 h every 2–3 weeks, or <0.5% of winter[27]) and may leave these groups and contact individuals in other groups or areas of the environment[14]. North American hibernating bats have recently been affected by white-nose syndrome, a disease caused by the fungal pathogen *Pseudogymnoascus destructans*[13,28]. We hypothesized that cryptic connections among species and individuals may be important in the transmission dynamics of white-nose syndrome, because the spread of the pathogen that causes this syndrome occurs rapidly within multiple species and some solitarily roosting species have suffered steep declines[14]. More generally, understanding the contribution of contacts outside social groups in pathogen transmission remains an important question in epidemiology.

We characterized connections among individuals using three methods to compare networks with and without cryptic contacts. We first measured direct physical contacts among bats, as well as links through membership of shared groups (a group of bats hibernating together), at eight sites at which bat communities were comprised of four species (Supplementary Table 1). For both measures, we included the additional connections bats made when switching between groups over the winter during periodic arousals from hibernation (see Methods). We simultaneously quantified the importance of cryptic connections in these communities by tracing epidemics of a surrogate pathogen, ultraviolet-fluorescent (UVF) dust (Extended Data Fig. 1), through each population. We estimated connections through physical contact, shared groups and UVF dust as the percentage of each species at each site that a given focal bat was connected to over the winter. By replicating epidemics within each population, we were able to account for variation in epidemic outcomes due to differences among individuals and species[29]. The fungus that causes white-nose syndrome subsequently invaded all eight sites at which we studied UVF-dust epidemics (Extended Data Fig. 2a–h). This enabled us to compare the transmission of *P. destructans* to connections within each community, as measured through UVF dust and shared group connections.

For hibernating bats, networks created from physical contact and shared groups substantially underestimated connections within populations, compared to epidemic networks of UVF dust (Fig. 1, Extended Data Fig. 3). Connections within species revealed by the

---

[1]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, USA. [2]Department of Biological Sciences, Virginia Polytechnic Institute, Blacksburg, VA, USA. [3]Wisconsin Department of Natural Resources, Bureau of Natural Heritage Conservation, Madison, WI, USA. [4]Department of Biology, Eastern Michigan University, Ypsilanti, MI, USA. [5]Michigan Department of Natural Resources, Baraga, MI, USA. [6]Michigan Department of Natural Resources, Norway, MI, USA. [7]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, NH, USA. [8]Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ, USA. [9]Bat Conservation International, Austin, TX, USA. *e-mail: hoytjosephr@gmail.com
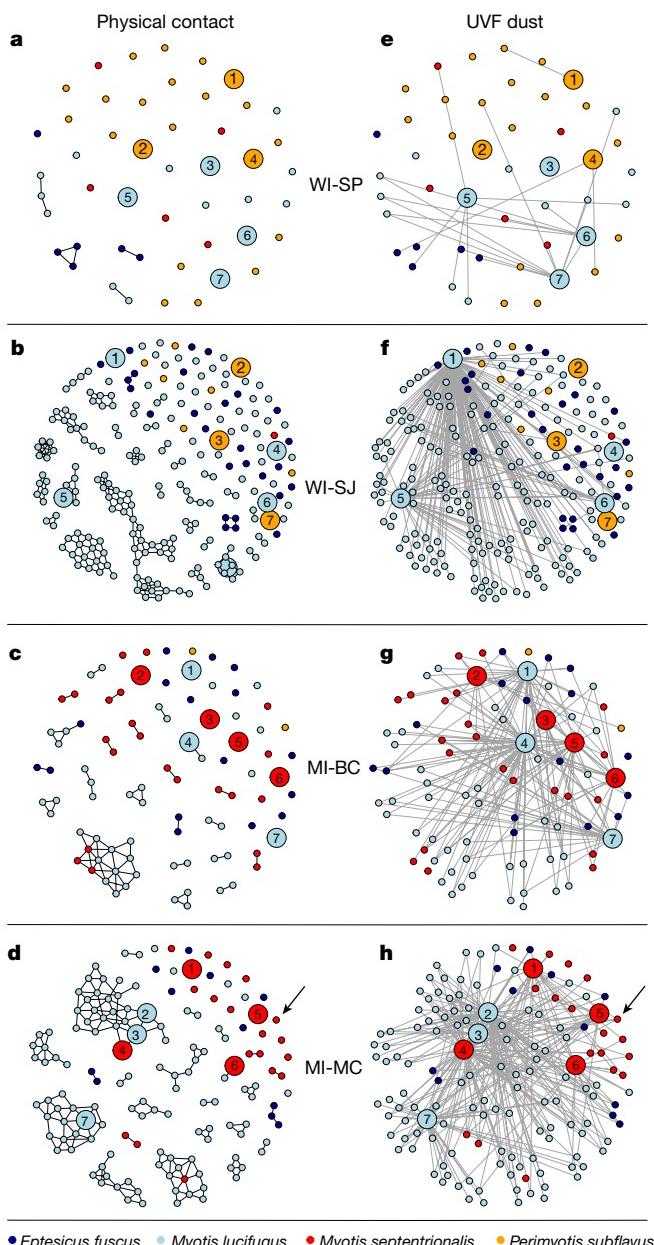
Fig. 1 | Observed physical contact and surrogate-pathogen epidemic networks for four communities of hibernating bats. a–h, Each row shows two networks—one based on physical contact and one based on the spread of UVF dust—for one of four sites (labelled WI-SP, WI-SJ, MI-BC and MI-MC). Each circle (node) represents an individual bat, and colour indicates species. Larger numbered circles (1–7) are bats that were UVF-dusted with unique colours in early winter. Lines (edges) between nodes indicate physical contact among bats in shared groups at the end of winter (**a–d**) or UVF dust that originated from a UVF-dusted bat (large circle) in epidemic networks (**e–h**). Each UVF-dust epidemic network in **e–h** shows seven simultaneous replicate epidemics, with each epidemic originating from a single, numbered UVF-dusted bat. For example, in **d** and **h** the arrow indicates a *M. septentrionalis* bat roosting solitarily (**d**) that was re-sighted with UVF dust from four different epidemics (from bats 1, 3, 4 and 7; **h**). Locations of nodes (bats) in the two networks are identical, so the spread of UVF dust within and between clusters of touching bats can be visualized in the UVF-dust networks.

transmission of UVF dust were threefold higher among individuals than was apparent from direct physical contact or shared groups for two of the three focal species (*Myotis lucifugus* and *Myotis septentrionalis*; Fig. 2, Extended Data Fig. 4, Supplementary Tables 2–4). Over 92% of *M. septentrionalis* individuals roosted alone (average group size of

1.34 ± 0.068 individuals; range 1–3; Extended Data Fig. 5) and were connected to only 7.3 ± 0.2% of other *M. septentrionalis* individuals by physical contact and 8.1 ± 0.2% by shared groups (Fig. 2b, left and middle, red columns; Supplementary Table 5, line 19). Despite the solitary roosting behaviour of this species, each *M. septentrionalis* individual was connected to 24.7 ± 4.3% of other *M. septentrionalis* individuals at each site by UVF dust (Fig. 2b, right, red column; Supplementary Table 5, lines 26, 27). For the most gregarious species—*M. lucifugus* (average group size 3.25 ± 0.34 individuals, range 1–53; Fig. 1, Extended Data Figs. 3, 5)—individuals were only in direct physical contact with 5.0 ± 0.1%, and in shared groups with 20.0 ± 0.2%, of the *M. lucifugus* population at each site (Fig. 2a, left and middle, light-blue columns; Supplementary Table 6, line 12), whereas UVF dust originating from each individual bat was transmitted to 28.6 ± 5.5% of the *M. lucifugus* individuals at each site (Fig. 2a, right, light-blue starred column, Supplementary Table 6, lines 1, 2). By contrast, for *Perimyotis subflavus*—another solitary species (99% of individuals roosted individually, $n = 62$; Fig. 1, Extended Data Figs. 3, 6)—there were few additional connections among individuals as detected by the UVF dust, when compared with physical contact or shared groups (Fig. 2c, orange columns; Supplementary Table 7).

The extent of environmental contamination with UVF dust for *P. subflavus* (533 ± 163 mm²) was not substantially lower than for the other species (929 ± 245 mm² and 1,100 ± 170 mm² for *M. lucifugus* and *M. septentrionalis*, respectively) and therefore was not sufficient to explain differences in connections (Extended Data Fig. 7a). However, the spatial overlap of UVF-dust colours in the environment was significantly lower for *P. subflavus* individuals (63.3 ± 5.8%) compared to *M. septentrionalis* and *M. lucifugus* (86.2 ± 3.7% and 86.8 ± 3.7%, respectively; Extended Data Fig. 7b), which suggests that spatial segregation may limit both indirect and direct transmission for *P. subflavus*.

Connections between species were underestimated even more by measurements of physical contact and shared groups than were connections within species (for example, see Fig. 2a). Overall, the spillover of UVF dust among species was 17- and 12-fold higher than expected based on physical contact and shared groups, respectively (Fig. 2). For example, *M. lucifugus* was—on average—in direct contact with only 0.0–0.1% of the *M. septentrionalis*, *P. subflavus* or *Eptesicus fuscus* populations (Fig. 2b; Supplementary Table 6, lines 14, 17, 20), and was only in shared groups with 0.0–1.7% of individuals for these three species (Fig. 2b; Supplementary Table 6, lines 24, 27, 30). However, inclusion of cryptic connections revealed through UVF-dust epidemics showed that each individual *M. lucifugus* was connected to an average of 22% of *E. fuscus*, 14% of *M. septentrionalis* and 1% of *P. subflavus* individuals at each site (Figs. 1, 2, Extended Data Fig. 3; Supplementary Tables 2, 6, lines 3, 6, 9).

For *M. lucifugus*, group size affected among-species and within-species transmission differently. The probability of an *M. lucifugus* individual having UVF dust from another *M. lucifugus* individual increased with group size, whereas UVF dust from *M. septentrionalis* was more likely to be found on solitary-roosting *M. lucifugus* individuals (Extended Data Fig. 8). This suggests that transmission of UVF dust within a site was density-dependent for *M. lucifugus*, and that *M. septentrionalis*—which are solitary—preferentially transmit UVF dust to solitary-roosting *M. lucifugus*.

The UVF-dust epidemics suggested that *P. destructans*, the fungus that causes white-nose syndrome, would spread rapidly through populations of some species of bats but not others (Fig. 2), and the subsequent invasion of *P. destructans* supported this prediction. Prevalence of *P. destructans* on *M. lucifugus* (Fig. 3a) and *M. septentrionalis* individuals (Fig. 3b) increased rapidly over the winter hibernation period (mean 84%, range 55–100%) at all sites. By contrast, transmission was much lower in *P. subflavus*; prevalence at three of the four sites remained below 25% at the end of winter (Fig. 3c). Across all species, UVF-dust transmission explained 3.5 times as much variation in *P. destructans* transmission as connections estimated using shared groups, and the similarity in transmission patterns between *P. destructans* and UVF
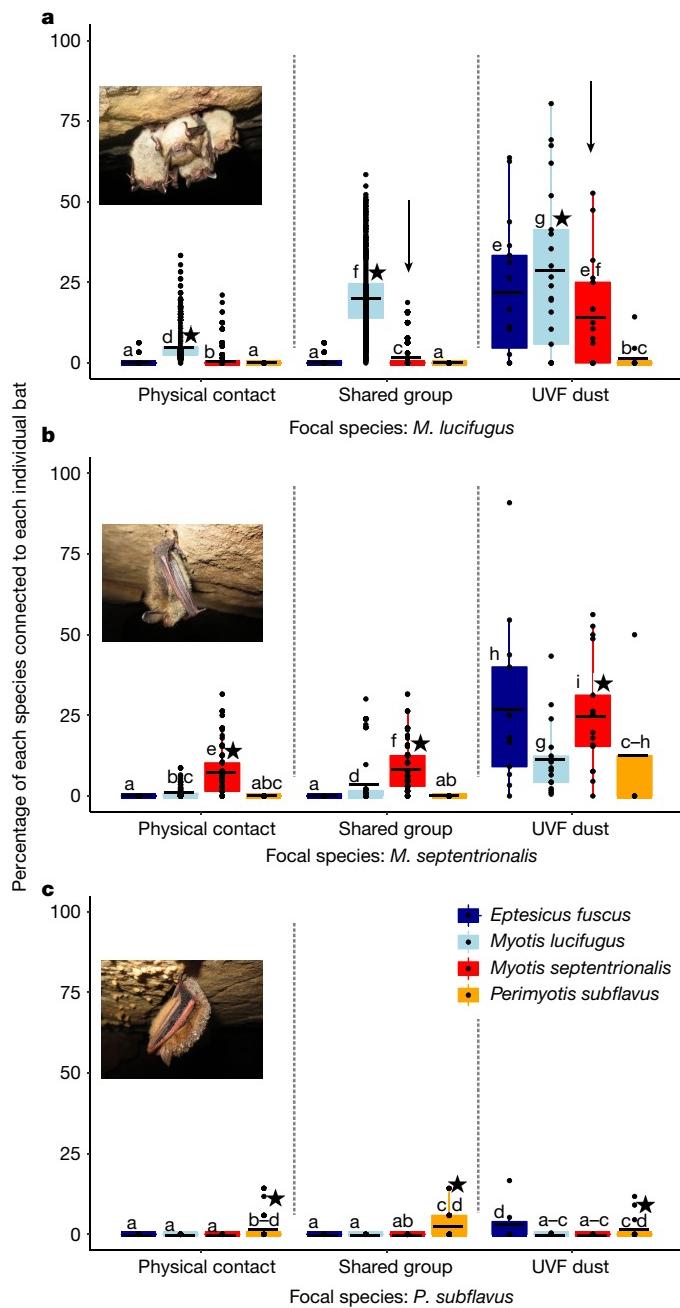
Fig. 2 | Contact, group and UVF-dust epidemic connections among individuals. Per cent of individuals of each species at each site observed in physical contact, in a shared group or with UVF dust that originated from a single focal individual. **a**, *M. lucifugus* ($n = 3,362$ physical contact estimates, $n = 3,345$ shared group estimates, $n = 74$ UVF-dust estimates). **b**, *M. septentrionalis* ($n = 413$ physical contact estimates, $n = 426$ shared group estimates, $n = 51$ UVF-dust estimates). **c**, *P. subflavus* ($n = 204$ physical contact estimates, $n = 199$ shared group estimates, $n = 45$ UVF-dust estimates). Each point indicates the percentage of the population of a single species at a site that is connected to a single focal individual. For example, the arrow in the right section of **a** shows the percentage of *M. septentrionalis* individuals at each site with each colour of UVF dust that originated from a UVF-dusted *M. lucifugus*. The arrow in the middle section of **a** shows the percentage of *M. septentrionalis* individuals at each site that were in a shared group with each individual *M. lucifugus* over the winter. Connection estimates for physical contact and shared social groups include connections that the bats made when switching positions in the network during arousals (see Methods). Letters above bars indicate groups for which 95% high posterior density intervals for pairwise differences did not include 0 (Supplementary Tables 5–7), and stars indicate within-species connections. The lower and upper hinges of the box plots show the first and third quartiles, and the black lines indicate the mean. The upper and lower whiskers extend to the largest and smallest value 1.5 times the interquartile range.
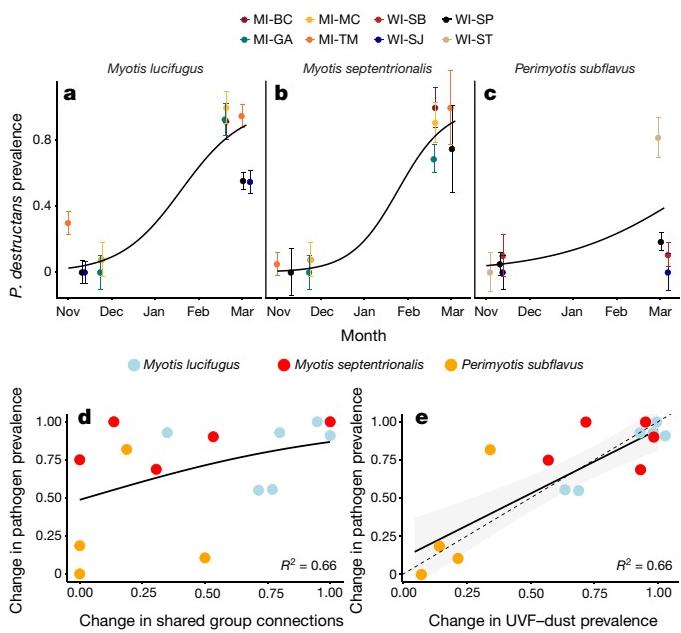
Fig. 3 | Pathogen and surrogate-pathogen transmission dynamics. Pathogen (*P. destructans*) prevalence over time during the first year of invasion, at the eight sites at which UVF-dust epidemics were studied for three species. **a**, *M. lucifugus* ($n = 207$ bats sampled). **b**, *M. septentrionalis* ($n = 116$ bats sampled). **c**, *P. subflavus* ($n = 125$ bats sampled). Points show pathogen prevalence (± 1 s.e.m.) for a species at a site and lines show fitted models for the change in pathogen prevalence over time; generalized linear mixed-effects model, binomial distribution with logit link and site as a random effect (**a**, *M. lucifugus*, $-3.71\ (\pm 0.80) + 1.43\ (\pm 0.20) \times$ month, $P = 3.14 \times 10^{-13}$; **b**, *M. septentrionalis*, $-1.14\ (\pm 0.20) + 0.41\ (\pm 0.27) \times$ month, $P = 1.11 \times 10^{-12}$; and **c**, *P. subflavus*, $0.50\ (\pm 1.18) + -0.75\ (\pm 0.30) \times$ month, $P = 0.003$). **d**, Pathogen transmission (transmission $= (\text{Pd}_{\text{March}} - \text{Pd}_{\text{November}})/(1 - \text{Pd}_{\text{November}})$, in which Pd is *P. destructans* prevalence) plotted against the change in shared-group connections, including movement between groups over winter ($n = 15$ species–site estimates). The curves show the nonlinear-mixed-model fitted relationship, with site as a random effect (pathogen transmission $= 1/(1 + e^{(-1.93\ (\pm 1.32) \times (\text{change in shared-group connections} - 0.03\ (\pm 0.29))))}$); $P = 0.19$). **e**, Pathogen transmission plotted against UVF-dust transmission ($n = 15$ species–site estimates). The line shows the fitted linear-mixed-model relationship, with site as a random effect, and the grey region shows the 95% confidence interval of the regression (pathogen transmission $= 0.13\ (\pm 0.12) + 0.825\ (\pm 0.15) \times$ UVF-dust transmission, $P = 0.0006$). The dashed line shows the 1:1 line for comparison. All P values are two-tailed.

dust demonstrates the importance of cryptic connections in driving epidemics (Fig. 3d, e).

Observing cryptic connections among individuals and groups via UVF-dust epidemics provides insight into the transmission patterns of *P. destructans*, which are difficult to understand from patterns of contact among species. *M. septentrionalis* and *P. subflavus* were both primarily observed as solitary-roosting individuals, which should limit transmission in both species relative to *M. lucifugus*, which roosts in larger groups (Extended Data Fig. 5). Although transmission rates in *P. subflavus* were indeed lower, transmission of *P. destructans* was equally high in *M. septentrionalis* and *M. lucifugus* (Fig. 3a, b). The UVF-dust epidemics suggested that this was because extensive cryptic connections link *M. septentrionalis* individuals with each other and to *M. lucifugus*. By contrast, *P. subflavus* had few observable or cryptic connections, and the lack of connections resulted in much lower rates of transmission. These results show that cryptic connections were not simply proportional to observed contacts, but vary among species and individuals. Cryptic connections uncovered by these data can also inform conservation actions that aim to reduce the effect of white-nose syndrome on bat populations (see 'Conservation implications' in Supplementary Information).

Characterizing all connections among individuals is vital to understanding, predicting and controlling outbreaks of infectious diseases[3,6,30]. We have shown that cryptic connections among groups and species have a key role in transmission dynamics and explain why transmission is far more intense in some species than others. Cryptic connections not only link social groups within species but also create bridges among species, resulting in highly connected communities and explosive epidemics. Characterizing the locations, frequencies and variation in cryptic connections is needed to more accurately predict and prevent future epidemics.

Online content

1. Woolhouse, M. E. J. et al. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342 (1997).
2. Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151 (2004).
3. Edmunds, W. J., O'Callaghan, C. J. & Nokes, D. J. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc. R. Soc. Lond. B* **264**, 949–957 (1997).
4. Snow, J. *On the Mode of Communication of Cholera* (John Churchill, London, 1855).
5. Salathé, M. & Jones, J. H. Dynamics and control of diseases in networks with community structure. *PLOS Comput. Biol.* **6**, e1000736 (2010).
6. Mossong, J. et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).
7. Nunn, C. L. & Altizer, S. *Infectious Diseases in Primates: Behavior, Ecology, and Evolution* (Oxford Univ. Press, Oxford, 2006).
8. Keeling, M. The implications of network structure for epidemic dynamics. *Theor. Popul. Biol.* **67**, 1–8 (2005).
9. Altizer, S. et al. Social organization and parasite risk in mammals: integrating theory and empirical studies. *Annu. Rev. Ecol. Evol. Syst.* **34**, 517–547 (2003).
10. Sah, P., Leu, S. T., Cross, P. C., Hudson, P. J. & Bansal, S. Unraveling the disease consequences and mechanisms of modular structure in animal social networks. *Proc. Natl Acad. Sci. USA* **114**, 4165–4170 (2017).
11. King, A. A., Ionides, E. L., Pascual, M. & Bouma, M. J. Inapparent infections and cholera dynamics. *Nature* **454**, 877–880 (2008).
12. Lloyd-Smith, J. O. et al. Epidemic dynamics at the human–animal interface. *Science* **326**, 1362–1367 (2009).
13. Lorch, J. M. et al. Experimental infection of bats with *Geomyces destructans* causes white-nose syndrome. *Nature* **480**, 376–378 (2011).
14. Langwig, K. E. et al. Sociality, density-dependence and microclimates determine the persistence of populations suffering from a novel fungal disease, white-nose syndrome. *Ecol. Lett.* **15**, 1050–1057 (2012).
15. Langwig, K. E. et al. Invasion dynamics of white-nose syndrome fungus, midwestern United States, 2012–2014. *Emerg. Infect. Dis.* **21**, 1023–1026 (2015).
16. Jones, K. E. et al. Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
17. Bansal, S., Grenfell, B. T. & Meyers, L. A. When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface* **4**, 879–891 (2007).
18. Gurley, E. S. et al. Person-to-person transmission of Nipah virus in a Bangladeshi community. *Emerg. Infect. Dis.* **13**, 1031–1037 (2007).
19. Alexander, K. A. et al. What factors might have led to the emergence of Ebola in West Africa? *PLoS Negl. Trop. Dis.* **9**, e0003652 (2015).
20. Weber, N. et al. Badger social networks correlate with tuberculosis infection. *Curr. Biol.* **23**, R915–R916 (2013).
21. Wendland, L. D. et al. Social behavior drives the dynamics of respiratory disease in threatened tortoises. *Ecology* **91**, 1257–1262 (2010).
22. McBryde, E. S. & McElwain, D. L. A mathematical model investigating the impact of an environmental reservoir on the prevalence and control of vancomycin-resistant enterococci. *J. Infect. Dis.* **193**, 1473–1474 (2006).
23. Poutanen, S. M. et al. Identification of severe acute respiratory syndrome in Canada. *N. Engl. J. Med.* **348**, 1995–2005 (2003).
24. Eubank, S. et al. Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004).
25. Lloyd-Smith, J. O. et al. Should we expect population thresholds for wildlife disease? *Trends Ecol. Evol.* **20**, 511–519 (2005).
26. Cross, P. C. et al. Integrating association data and disease dynamics in a social ungulate: bovine tuberculosis in African buffalo in the Kruger National Park. *Ann. Zool. Fenn.* **41**, 879–892 (2004).
27. Thomas, D. The physiological ecology of hibernation in vespertilionid bats. *Symp. Zool. Soc. Lond.* **67**, 233–244 (1995).
28. Langwig, K. E. et al. Drivers of variation in species impacts for a multi-host fungal disease of bats. *Phil. Trans. R. Soc. Lond. B* **371**, 20150456 (2016).
29. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
30. Edmunds, W. J., Kafatos, G., Wallinga, J. & Mossong, J. R. Mixing patterns and the spread of close-contact infectious diseases. *Emerg. Themes Epidemiol.* **3**, 10 (2006).

METHODS

Study sites and sampling for *P. destructans*. We studied contact rates, social networks and fungal infection with *P. destructans* in bats at eight abandoned mines in Wisconsin and Michigan over five winters (Extended Data Fig. 2a–h, Supplementary Table 1; winters of 2012/2013–2016/2017). Sampling to measure pathogen transmission was conducted over all five winters and surrogate-pathogen tracing (UVF dust) occurred over two winters (2013/2014 and 2014/2015; Extended Data Fig. 2a–h). There were 183 ± 72 (range 18–624) total bats of three or four species hibernating at each site (Supplementary Table 1). Bats in this region begin hibernating in September–October and leave hibernacula in April–May. We visited each site twice during each winter in all five years. In November and March, we counted all bats by species and sampled bats to test for the presence of *P. destructans*, the fungus that causes white-nose syndrome (Extended Data Fig. 2a–h). We sampled up to 20 individuals of each species for *P. destructans*, using a previously described swab-sampling technique during each visit[31,32]. Samples were placed in RNAlater (Thermo Fisher Scientific) and subsequently tested for *P. destructans* by qPCR[31,33]. A power analysis was used to predetermine sample size for pathogen data collection. At six sites, *P. destructans* invasion occurred after the UVF-dust study was conducted (1–3 winters later; Extended Data Fig. 2). At two sites, *P. destructans* was detected during the UVF-dust study (Extended Data Fig. 2b, c). However, there were no observable declines over the winter at these two sites and observable white-nose syndrome symptoms were largely absent in the first year, as has been observed at other sites[15].

Quantifying direct contact, shared groups and UVF-dust transmission. We quantified connections among hibernating bats in three ways: (1) bats that were physically touching each other, (2) bats hibernating in groups (a group of bats touching each other) and (3) bats sharing a surrogate pathogen (a unique colour of UVF dust). The first two measures of social interactions—physical contact and shared social groups—are commonly used to characterize human, livestock and wildlife communities[6,10]. Transmission of UVF dust mirrors the transmission ecology of *P. destructans*, which is an epidermal fungal pathogen that is transmitted by direct or indirect contact among bats and the environment.

In November, we applied a unique colour of UVF dust randomly to each of 4–7 individual male bats at each site (with one exception; see below). To determine potential sample sizes for field experiments, we assessed the ability of four independent observers to distinguish and re-sight colours of UVF dust. Both large patches and trace amounts of dust were spread on rock surfaces and then examined to determine which colours could be distinguished from each other. We initially compared 10 colours of UVF dust and found that with a combination of UV and visible light, seven could be reliably distinguished from each other. The seven unique colours included in the study were: ECO-11 Aurora Pink, ECO-15 Blaze Orange, 16 Arc Yellow, ECO-17 Saturn Yellow, ECO-18 Signal Green, ECO-19 Horizon Blue (Day-Glo Colour); and DFSB-C0 Clear Blue (Risk Reactor).

We randomly applied one gram of a single colour of UVF dust to each individual bat. Each bat was dusted over a clean bag to ensure that there was no transfer of UVF-dust colours between bats (Extended Data Fig. 1a). The UVF dust was applied by spreading the dust along the entire dorsal and ventral surface of the bat, and limiting the application of dust to the face and head (Extended Data Fig. 1b, c). We used the same quantity of dust for all three species because they broadly overlap in body size (*M. lucifugus*: 222–269 mm (wing span), 60–102 mm (length); *M. septentrionalis*: 230–260 mm (wing span), 78 mm (length); and *P. subflavus*; 220–250 mm (wing span), 77–89 mm (length)[34–36]). We also placed a unique wing band on each bat for identification.

We returned to each site during March and searched the environment and inspected all bats for UVF dust using UV flashlights (395 nm; Hayward) and visible light. The high reflectance of the UVF dust made dusted areas in the environment and on bats easily visible. We carried a colour key that consisted of small amounts of the seven colours of UVF dust fixed to black construction paper to help to confirm colour identification. Each colour of UVF dust observed on each bat was recorded (Extended Data Fig. 1d, e) and observers were blinded during resighting to the original species colour combination at each site. To aid in identification of the colour of small patches of dust, a piece of clear tape was used to collect a sample of the dust. This tape was affixed to a plastic microscope slide, and analysed under a dissecting microscope in the laboratory to confirm the colour. The number of individuals detected with each UVF-dust colour was used as an estimate of the total epidemic size resulting from a single (dusted) individual and includes both direct bat-to-bat contacts in and outside shared hibernating groups, and indirect bat–environment–bat contacts.

We also quantified the area of each patch of UVF dust, by colour, in the hibernacula environment to the nearest square centimetre (Extended Data Fig. 1f) and mapped all environmental contacts to the nearest metre location using a metre tape. For sites with branching passages or passages wider than 2 m, we indicated the location of each environmental contact on a map of the hibernaculum. A team of 2–3 people visually inspected the entire surface area of the site using UV

flashlights. Each patch of UVF dust (Extended Data Fig. 1f) was measured using a fine-scale (mm) ruler to estimate the size of each patch of UVF dust. If there was any uncertainty in the colour of a patch of dust, we collected a sample of the dust as described above.

We dusted individuals of three species of bats, *M. lucifugus*, *M. septentrionalis* and *P. subflavus*. We refer to *M. lucifugus* as the more social species because it often (>50%) occurred in groups of more than one bat. By contrast, over 90% of individuals of the two other species roosted alone, and thus we refer to them as solitary. A fourth species (*E. fuscus*) was present at six of the eight sites. We did not dust individuals of this species, but we did inspect them for UVF dust. At six sites, we dusted three individuals of one species and four individuals of another (Supplementary Table 1). At one site, we only dusted four individuals of *P. subflavus* because the abundance of other species was too low (Supplementary Table 1). At an eighth site, we dusted five *M. septentrionalis* with five unique colours and five *M. lucifugus* with a single colour. For analyses for this site, we divided the total number of individuals receiving dust from *M. lucifugus* by five and included this estimate as a single data point in relevant analyses (Fig. 2, Supplementary Table 1).

To quantify direct connections among bats, we recorded connections of bats in physical contact and in shared groups in March. Bats spend ~99.3% of winter in these groups, making them similar to social groups in other systems[27]. However, bats do sometimes switch among groups during winter. To quantify direct connections among bats via shared groups, we performed simulations to estimate the number of direct contacts and shared connections when the bats periodically aroused from hibernation in between our visits to the site in November and March. We used several datasets to examine variation in group sizes, cluster mate fidelity within a winter and UVF-dust transmissibility to help to guide simulations.

First, we investigated whether the group size (clustering behaviour) of hibernating bats differed over the winter period. We analysed data collected over a five-year period on the group sizes of the four species of bats mentioned above. We compared social-group size between early and late winter using generalized linear mixed-effects models with a Poisson distribution and a log-link. We found no significant difference in behaviour, range and frequency of group size, over the winter across the three UVF-dusted species (Extended Data Fig. 6a). Group sizes of *E. fuscus* slightly increased over the winter, but the probability of roosting with other species did not change (logistic regression with site as a random effect: early winter, intercept: -3.23 ± 0.59; late winter, coefficient -0.66 ± 0.93, $P = 0.476$) and within-species transmission was not estimated for *E. fuscus*, because we did not apply UVF dust to this species. Our analyses of group-size data also indicated that there were no significant differences in roosting behaviour between the two winters during which the UVF-dust data were collected and the winters during which the pathogen data were collected (Extended Data Fig. 6b), which suggests that group sizes were relatively consistent over time before populations declined from white-nose syndrome[14].

Second, we examined the probability that bats were found in groups with the same bats in November and March using data from individually banded bats observed within a single winter. These data indicate that for 24 pairs of bats that roosted in a group together in November, only 2 pairs were found together in March (despite both members of the pair being observed in both November and March). Nine of these bats became solitary between November and March, and ten individuals roosting solitarily in early winter were found in shared groups in March. These data suggest that only some bats remain with the same roost-mates over winter, and that many bats re-assort among groups during hibernation.

Finally, we conducted an experiment to assess the number of transmission chains through which UVF dust spreads (for example, primary, secondary transmission and so on). We initially captured 18 *M. lucifugus* individuals. We dusted three bats and placed them individually in $38 \times 38 \times 61$-m mesh terrariums (Restcloud, via Amazon). After one hour, we added a single individual male *M. lucifugus* to each enclosure with a single dusted bat to examine transmission from a dusted bat to a primary contact. After 60 s, we removed the primary-contacted bats and inspected them for UVF dust, as described above. We then placed the primary-contacted bats individually into clean cages. After 60 min, we placed one non-dusted individual male bat into each of the enclosures with each primary contacted bat for 60 s, to examine secondary transmission. We repeated this process for dusted bats, primary bats and secondary bats four hours later, for a total of nine measurements of transmission between primary-contacted bats and secondary-contacted bats, and six measurements of transmission between originally dusted bats and primary bats. We found that all of our primary contacted bats (6 out of 6) had UVF dust immediately following interaction with the originally UVF-dusted bats, and this dust remained visible for the duration of the experiment (6 h). Three of the nine secondary-contacted bats initially had UVF dust after their interactions with the primary-contacted bats, but none of these three bats retained the UVF dust 2–5 h later (Supplementary Table 8; UVF dust was probably removed by bats grooming the small amounts of dust they obtained

during secondary contact). This experiment suggests that UVF dust observed on bats in March most probably came from primary contacts between originally dusted bats or from the environment, and not from secondary or tertiary contacts among individuals.

We used these data to guide simulations to estimate the number of bats at each site with which a focal bat was likely to have shared direct physical and social-group connections, between our visits to the site. We used the observed contact networks of bats observed in March (Fig. 1, Extended Data Fig. 3), and randomly rearranged bats within each site while maintaining the same distribution of connections for each rearrangement using R packages 'igraph'(http://igraph.org) and 'picante'[37], keeping track of all primary connections. We performed six rearrangements, which reflects the average arousal frequency of hibernating bats (16 days[38,39]) and the duration between our visits to examine the spread of the UVF dust (90 days; range: 88–93 days), resulting in approximately 5.6 rearrangements over our study. Each site adjacency matrix was divided into within- and among-species blocks for each species combination (M. lucifugus–M. lucifugus (block 1), M. lucifugus–M. septentrionalis (block 2) and so on). We randomized within each block six times, maintaining the total number of edges or connections for each combination from the original social network. Symmetry was maintained for within-species connections in the matrix. We allowed each individual to contact the same individuals or return to the same position in the network. We then calculated the total number of unique direct connections per individual by summing across all six randomly generated matrices or rearrangements.

We performed simulations for shared-group rearrangement in a similar manner. Individuals were randomly rearranged among positions held by the same species within the network, while maintaining the edge distribution within blocks and including the possibility of an individual remaining a member of the same group. We summed the unique number of individuals that a focal individual was connected to over the six rearrangements.

We divided the total number of connections for each individual, made through physical contacts and social-group sharing, by the total population of that species at a site to calculate the percentage of each species connected to a focal individual. The percentages represent the total direct connections made by bats over the winter. A comparison of connections from the snapshot data to connections estimated using six rearrangements is provided in Extended Data Fig. 4.

Analyses. We quantified the percentage of bats of each species at a site that were connected to each focal bat either through physical contact, shared social group membership or sharing a colour of UVF dust. For the connection estimates using the first two measures (physical contact and shared groups), every bat in the site served as a focal bat and all other bats were Bernoulli trials in a binomial sample. For UVF-dust connections, each bat originally dusted in November served as a focal bat, and all other bats were a Bernoulli trial in a binomial sample. We compared the number of connections among connection types (physical contact, shared group member and UVF dust) and species (including within- versus between-species) by fitting a binomial hierarchical model with Bayesian methods using the no-U-turn sampler, an extension of Hamiltonian Markov chain Monte Carlo, with site as a random effect. We fit models and estimated the posterior distributions for all parameters using the R package 'brms'[40], which uses the programing language Stan via the 'rstan' package (https://cran.r-project.org/web/packages/rstan/index.html). We included weakly informative priors for all parameters (normal distribution with mean of zero and standard deviation of five). We ran a total of 4 chains for 2,000 iterations each, with a burn-in period of 1,000 iterations per chain resulting in 4,000 posterior samples, which—given the more efficient no-U-turn sampler[40]—was sufficient to achieve adequate mixing and convergence. All $\hat{R}$ values were less than or equal to 1.01 indicating model convergence.

We assessed differences in connections between species and by connection type (direct contact, shared social group and UVF dust) by determining whether the 95% credible intervals for pairwise comparisons between any two groups included zero (for example, physical contact between M. lucifugus and P. subflavus and UVF-dust connections between M. lucifugus and M. septentrionalis). We used Bayesian methods in this analysis to account for issues of complete separation for some species and data-type combinations. For example, P. subflavus was never in physical contact with M. septentrionalis, and the parameter estimating the difference between this binomial probability in a logistic regression and one in which there are some connections or non-zero values (for example, M. lucifugus in contact with M. lucifugus) would be undefined in a frequentist analysis. The inclusion of (weak) prior information in the Bayesian approach makes it possible to handle these complete separation issues.

We assessed potential biases in re-sighting different colours of UVF dust by comparing models including both site and UVF-dust colour as random effects to models with only site as a random effect, using Bayesian methods identical to those described above. These models were only fit to the UVF-dust connection dataset because differences between UVF-dust colours do not apply to other connection types. The coefficient estimates and credible intervals for each species connection

were similar among analyses including and excluding colour as a random effect (Supplementary Table 9).

We also examined whether the probability of having UVF dust increased with group size for M. lucifugus. For each origin species (M. lucifugus, M. septentrionalis and P. subflavus), we examined correlations between M. lucifugus cluster size and the probability of having UVF dust from any dusted individual, using generalized linear mixed-effects models with a binomial distribution and a logit link with site and individual bat as random effects, and group size as a fixed effect, using the 'lme4' package[41]. Individual bat was included as a random effect because some individuals were re-sighted with multiple colours of UVF dust, and thus are replicated in the analyses. One site (MI-GA) was excluded from the within-species M. lucifugus analysis because five individuals were dusted with the same colour at this site.

We examined differences among species in the probability of becoming infected with P. destructans over the winter at the same sites, using a generalized linear mixed-effects model with a binomial distribution and logit link, with species interacting with date as fixed effects, and site as a random effect. We examined the correlation between P. destructans transmission and both UVF-dust transmission and shared-group connections. We regressed the change in P. destructans prevalence over the winter during the first year of fungal invasion on the change in UVF-dust prevalence over the winter, or the change in shared-group connections over the winter for the same dusted individuals at each site using linear and non-linear mixed-effects models with site as a random effect using the 'nlme' package[42].

We examined differences among species in the total surface area of the environment covered with UVF dust, using a linear mixed-effects model with species as a fixed effect, and site as a random effect. Finally, we examined the differences in spatial segregation within hibernacula for each bat originally dusted in November. The 'home range' or area within each hibernaculum used by an individual was calculated by using locations in the environment in which UVF dust from each of our originally dusted bats was detected, as described above. We estimated home range overlap for each focal bat by computing the proportion of the home range of one focal bat that is overlapped by another focal bat using the 'kerneloverlap' function in the R package 'adehabitatHR'[18]. This function uses the kernel utilization distribution, which is a bivariate function giving the probability density that an animal is found at a point according to its geographical coordinates. We removed 10% of the outlying points, and restricted analyses to individuals with >10 dust locations detected in each site[18]. To estimate a kernel utilization distribution for sites at which dust locations were taken with a single coordinate (length along the mine tunnel), we created an additional coordinate location (tunnel width) by drawing 100 times from a uniform distribution between 0 and 2 (the approximate width of the mine tunnel, 2 m), and averaging percentage overlap values for each individual. We then compared the percentage overlap of areas used by each species using a linear mixed-effects model with species as a fixed effect, and site as a random effect. All statistical tests were carried out using R 3.3.2.

We complied with all relevant ethical regulations and all work was approved and performed under protocol FrickW1106, approved by the University of California, Santa Cruz IACUC.

Code availability. Supporting code and the comma separated value data files are available via GitHub at https://github.com/hoytjosephr/cryptic-connections.git.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All raw data points are contained within the main-text Figs. and Extended Data Figs. All other data are available from the corresponding author upon reasonable request.

31. Langwig, K. E. et al. Host and pathogen ecology drive the seasonal dynamics of a fungal disease, white-nose syndrome. Proc. R. Soc. Lond. B **282**, 20142335 (2014).
32. Hunt, A., Collins-Poythress, J. & Langwig, K. E. Swabbing protocol for Geomyces destructans https://www.youtube.com/watch?v=KU1EJPJXNPk (2013).
33. Muller, L. K. et al. Bat white-nose syndrome: a real-time TaqMan polymerase chain reaction test targeting the intergenic spacer region of Geomyces destructans. Mycologia **105**, 253–259 (2013).
34. Fujita, M. S. & Kunz, T. H. Pipistrellus subflavus. Mamm. Species **228**, 1–6 (1984).
35. Dawson, D. L. Functional interpretations of the radiographic anatomy of the femora of Myotis lucifugus, Pipistrellus subflavus, and Blarina brevicauda. Am. J. Anat. **157**, 1–15 (1980).
36. Caceres, M. C. & Barclay, R. M. R. Myotis septentrionalis. Mamm. Species **634**, 1–4 (2000).
37. Kembel, S. W. et al. Picante: R tools for integrating phylogenies and ecology. Bioinformatics **26**, 1463–1464 (2010).
38. Warnecke, L. et al. Inoculation of bats with European Geomyces destructans supports the novel pathogen hypothesis for the origin of white-nose syndrome. Proc. Natl Acad. Sci. USA **109**, 6999–7003 (2012).

39. Czenze, Z. J., Jonasson, K. A. & Willis, C. K. R. Thrifty females, frisky males: winter energetics of hibernating bats from a cold climate. *Physiol. Biochem. Zool.* **90**, 502–511 (2017).

40. Bürkner, P.-C. brms: an R package for bayesian multilevel models using stan. *J. Stat. Softw.* **80**, 1–28 (2017).

41. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

42. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. nlme: linear and nonlinear mixed effects models. R package version 3.1-137 https://cran.r-project.org/web/packages/nlme/index.html (2018).
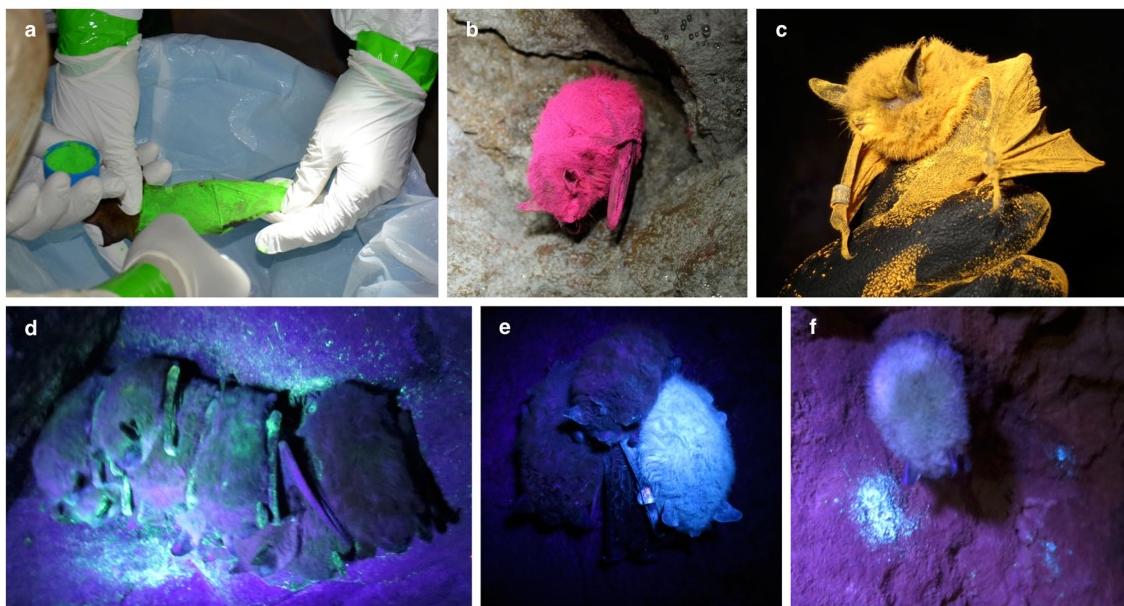
Extended Data Fig. 1 | Application and re-sighting of UVF dust on bats and in the environment. a, Applying UVF dust to *P. subflavus*. **b**, A dusted *P. subflavus* shortly after release. **c**, *M. septentrionalis* after UVF-dust application. **d**, A group of eight *M. lucifugus* roosting together in March. Each bat in the group had UVF dust. **e**, A dusted *M. lucifugus* in a group of three bats during March. At least three of the bats in the photo can be seen with UVF dust. **f**, Location in the environment with patches of UVF dust in March next to a hibernating *P. subflavus*.
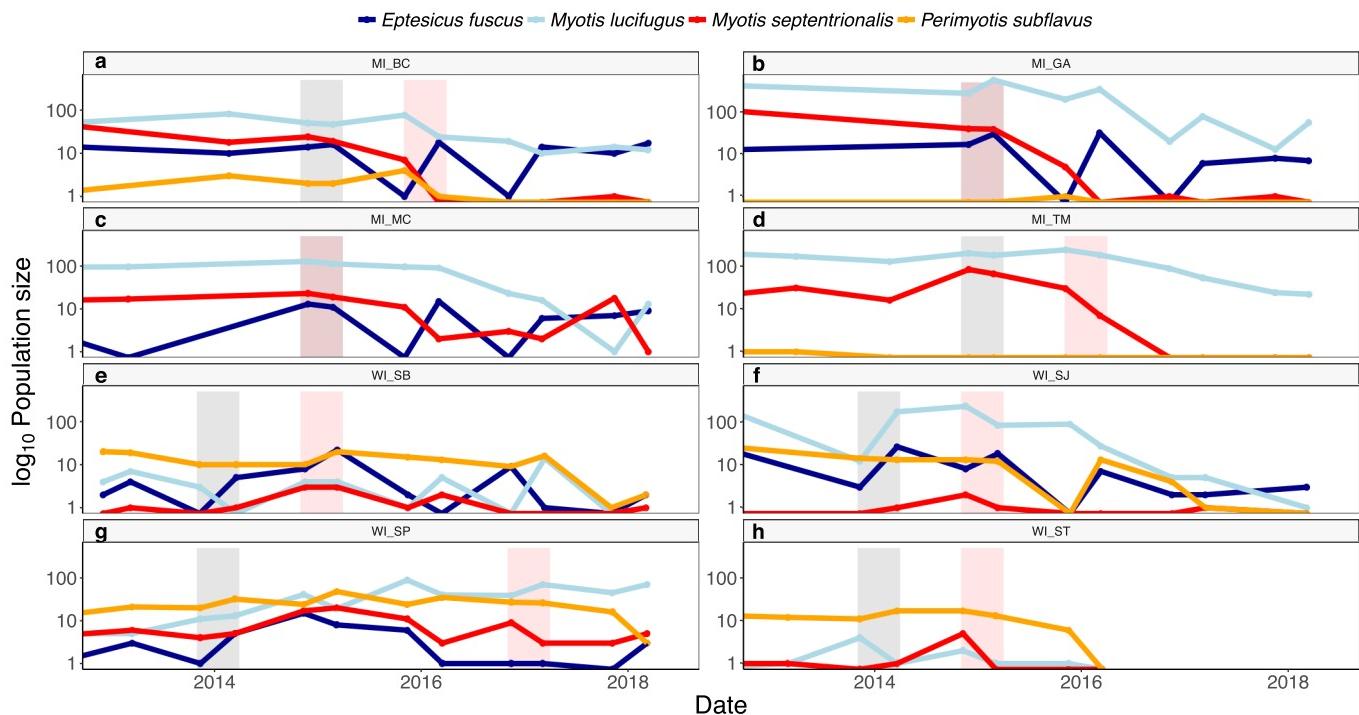
Extended Data Fig. 2 | Trends of population counts and timing of UVF-dust and disease data collection. The grey bar shows the winter period when the UVF-dust study was conducted and the red bar shows the year the *P. destructans* arrived. Temporal patterns of social group size are provided in Extended Data Fig. 6a.
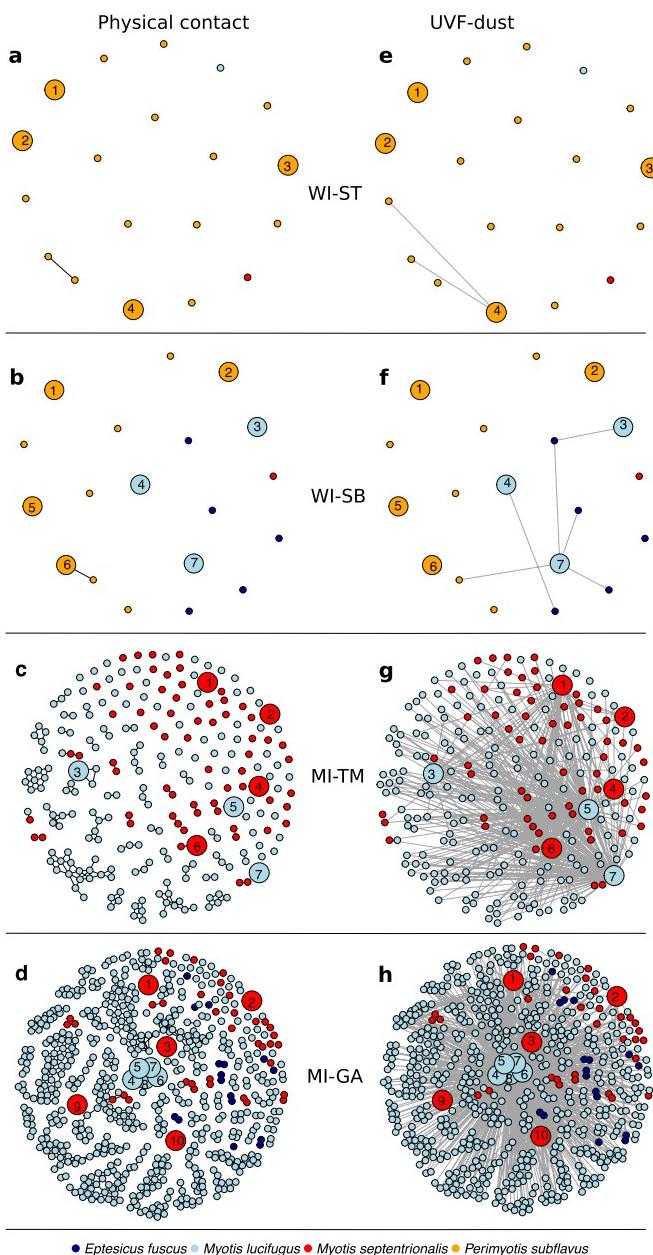
Extended Data Fig. 3 | Social and epidemic networks for four additional communities of hibernating bats. a–h, Each row shows two networks (physical contact and UVF dust) for each of four sites (WI-ST, WI-SB, MI-TM and MI-GA). Each circle (node) represents an individual bat, with colour indicating species; larger numbered circles (1–7) are bats that were dusted with UVF dust with unique colours in November at each site. Lines (edges) between nodes indicate physical contact among bats in social networks (**a–d**) or UVF dust that originated from the UVF-dusted bat (large circle) in epidemic networks (**e–h**). Each UVF-dust epidemic network in **e–h** shows seven replicate epidemics, with each epidemic originating from a single, numbered UVF-dusted bat. Locations of nodes (bats) in the two networks are identical, so the spread of UVF dust within and between groups of touching bats can be visualized in the UVF-dust network.
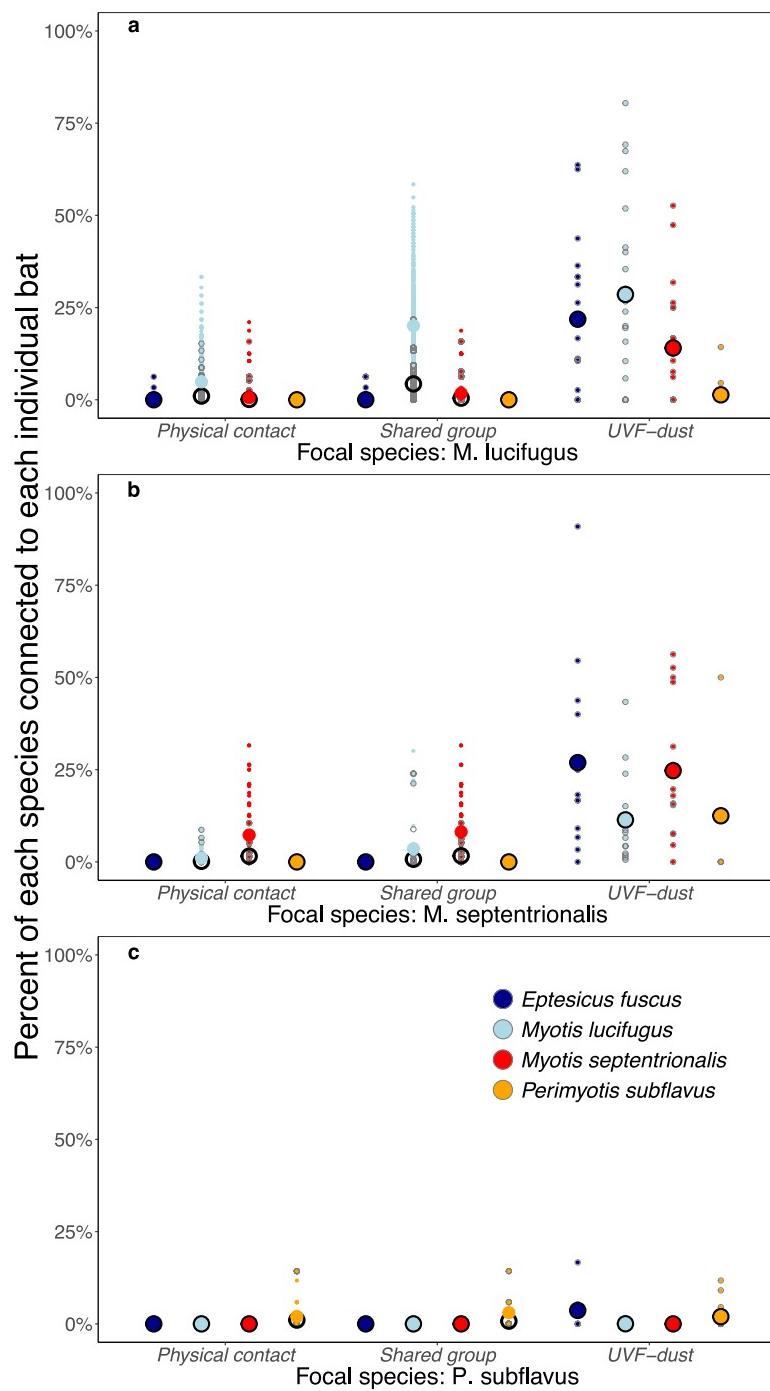
Extended Data Fig. 4 | Comparison of observed data from late-winter and total shared-group connections including group rearrangements. The *y* axes show the percentage of individuals of each species at each site observed in direct physical contact, in shared groups or with UVF dust originating from a single focal individual. **a**, *M. lucifugus*. **b**, *M. septentrionalis*. **c**, *P. subflavus*. Small points show the observed (grey) data and augmented (coloured) connections, including group rearrangements. The black open points show the mean late winter estimate of direct contact and social-group connections based on a single time point of the populations without group switching (Fig. 1, Extended Data Fig. 3; left panels). The large coloured points (means in Fig. 2) show the mean number of connections from simulations in which bats rearranged during six arousals over the winter (also shown in Fig. 2). Where the large coloured points are within the black open points, there were no differences between the observed and augmented data including group rearrangements. The connections observed in UVF-dust epidemics are shown for comparison, but do not differ.
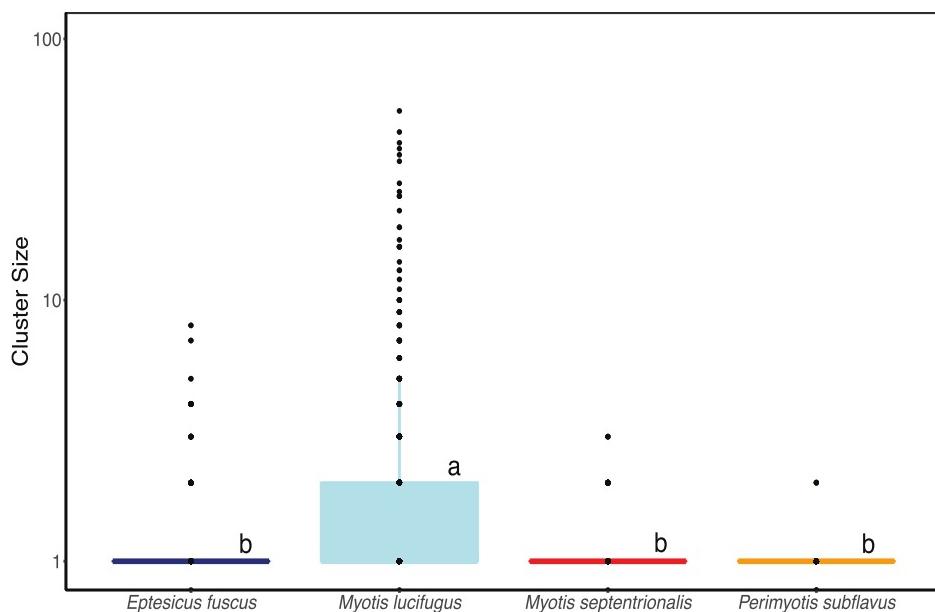
Extended Data Fig. 5 | Shared-group (cluster) size for four species of hibernating bats. Different letters above bars indicate mean group sizes that differ significantly; generalized linear model with negative binomial distribution with site as a random effect and species as a fixed effect (*E. fuscus*, $n = 202$ observed groups, intercept: 0.020 ± 0.11; *M. lucifugus*, $n = 1,011$ observed groups, coefficient 0.450 ± 0.08; *M. septentrionalis*, $n = 151$ observed groups, coefficient: -0.03 ± 0.12; *P. subflavus*, $n = 263$ observed groups, coefficient, -0.058 ± 0.11). The lower and upper hinges of the box plot show the first and third quartiles. The upper and lower whiskers extend to the largest and smallest value 1.5 times the interquartile range.
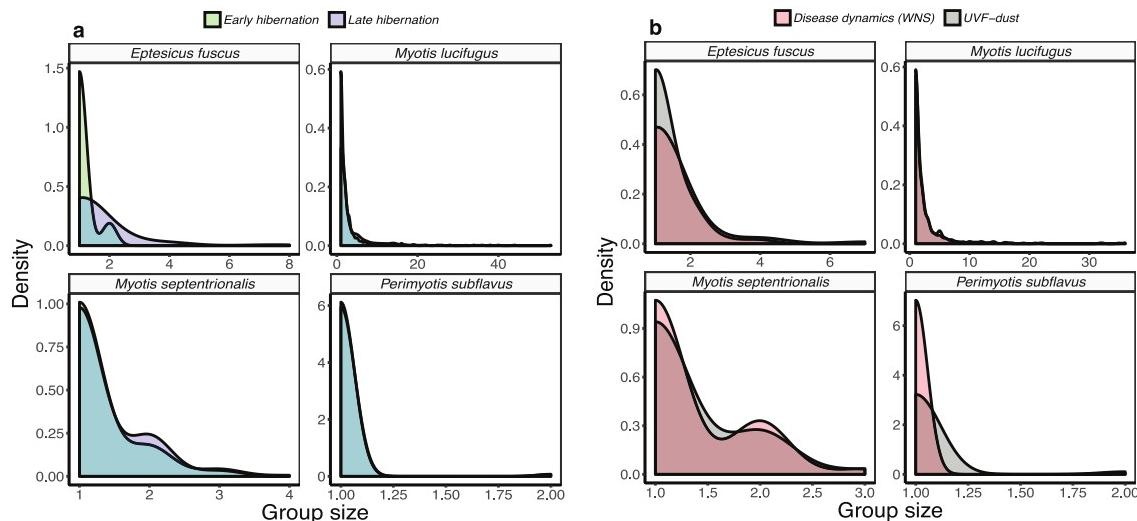
Extended Data Fig. 6 | Distributions of shared-group sizes for four species of hibernating bats in early (November) and late (March) winter and during the collection of UVF-dust and fungal infection data.
a, Across all species, there was no significant change in social-group size over the winter using season (early versus late winter) as a fixed effect, and site and species as random effects (early winter, intercept: 0.38 ± 0.21; late winter, coefficient: 0.01 ± 0.03, $P = 0.716$, $n = 1,956$ bat groupings). We also investigated changes in social-group size for *M. lucifugus*, using an identical model but dropping species as a random effect. Again, we found no significant change in social group size over the winter (early winter, intercept: 0.84 ± 0.17, $n = 457$; late winter, coefficient: -0.008 ± 0.04,

$n = 528$, $P = 0.84$). Neither model was significantly better than a null-intercept model. For *E. fuscus*, shared-group sizes increased slightly over the winter (early winter, intercept: 0.11 ± 0.11, $n = 79$; late winter, coefficient: 0.28 ± 0.14, $n = 100$, $P = 0.040$). **b**, We compared social-group size ($n = 1,328$) between the disease-arrival year and the dust-study year using a generalized linear mixed-effects model with a Poisson distribution and a log link. We found no significant difference in social-group sizes among years using study year as a categorical fixed effect (dust or disease) and site and species as random effects (disease study, intercept: 0.32 ± 0.16; UVF-dust study, coefficient: -0.04 ± 0.04, $P = 0.33$).
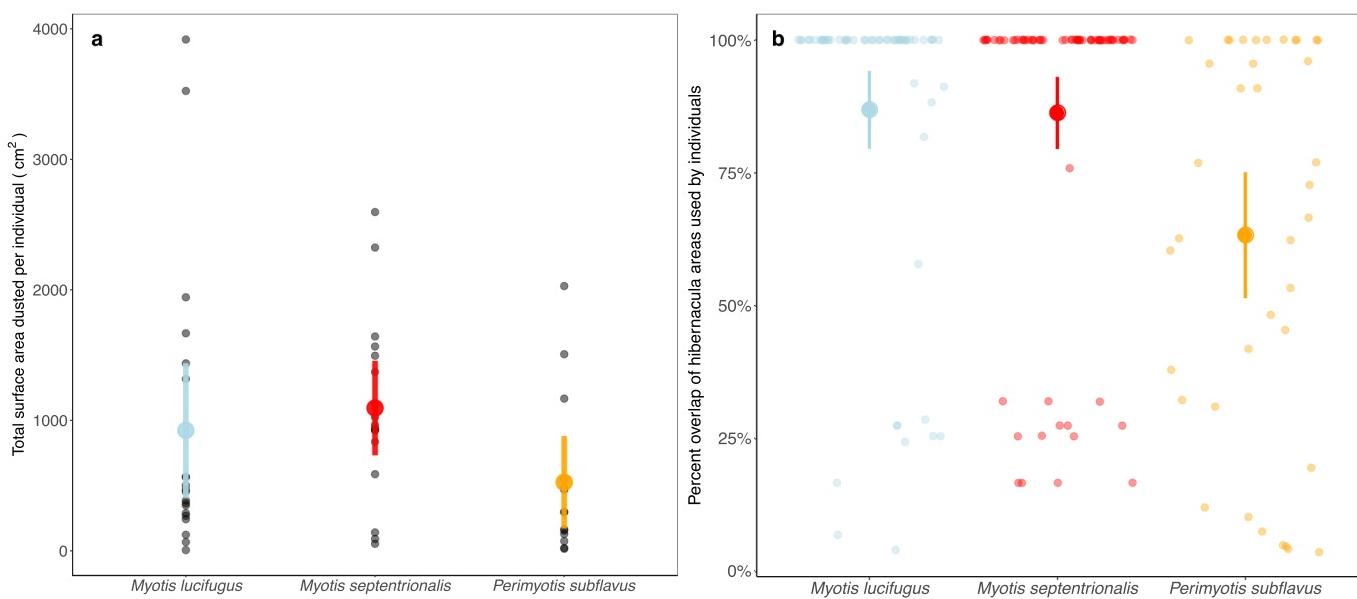
Extended Data Fig. 7 | Environmental contamination and spatial overlap of UVF-dusted individuals. a, Total surface area of the hibernacula walls and ceilings with UVF dust originating from different species ($n = 51$ measurements). Small black points show the summed total surface area with each colour of dust. Large points show the mean and 95% confidence intervals of the three species that were originally dusted in November (generalized linear mixed-effects model with site as a random effect and species as a fixed effect; *M. lucifugus*, intercept: 2.71 ± 0.15; *M. septentrionalis*, coefficient: 0.14 ± 0.17, $t = 0.79$; *P. subflavus*, coefficient: -0.21 ± 0.18, $t = -1.16$; species effect, $P = 0.30$). **b,** Percentage overlap of hibernacula areas used by individual bats compared to areas used by other individuals ($n = 172$ individual overlap estimates). The home range was calculated for each individual UVF-dust colour based on locations in the environment at which bats deposited their unique colour of UVF dust (see Methods). The points show an overlap estimate for each individual dusted bat compared to another dusted bat at that same site with larger points showing the mean and 95% confidence intervals for each species (generalized linear mixed-effects model with site as a random effect and species as a fixed effect; *M. lucifugus*, intercept: 0.88 ± 0.05; *M. septentrionalis*, coefficient: -0.03 ± 0.05, $t = -0.50$; *P. subflavus*, coefficient: -0.21 ± 0.07, $t = -2.86$; species effect, $P = 0.02$). Points are jittered in **b** to show overlapping data.
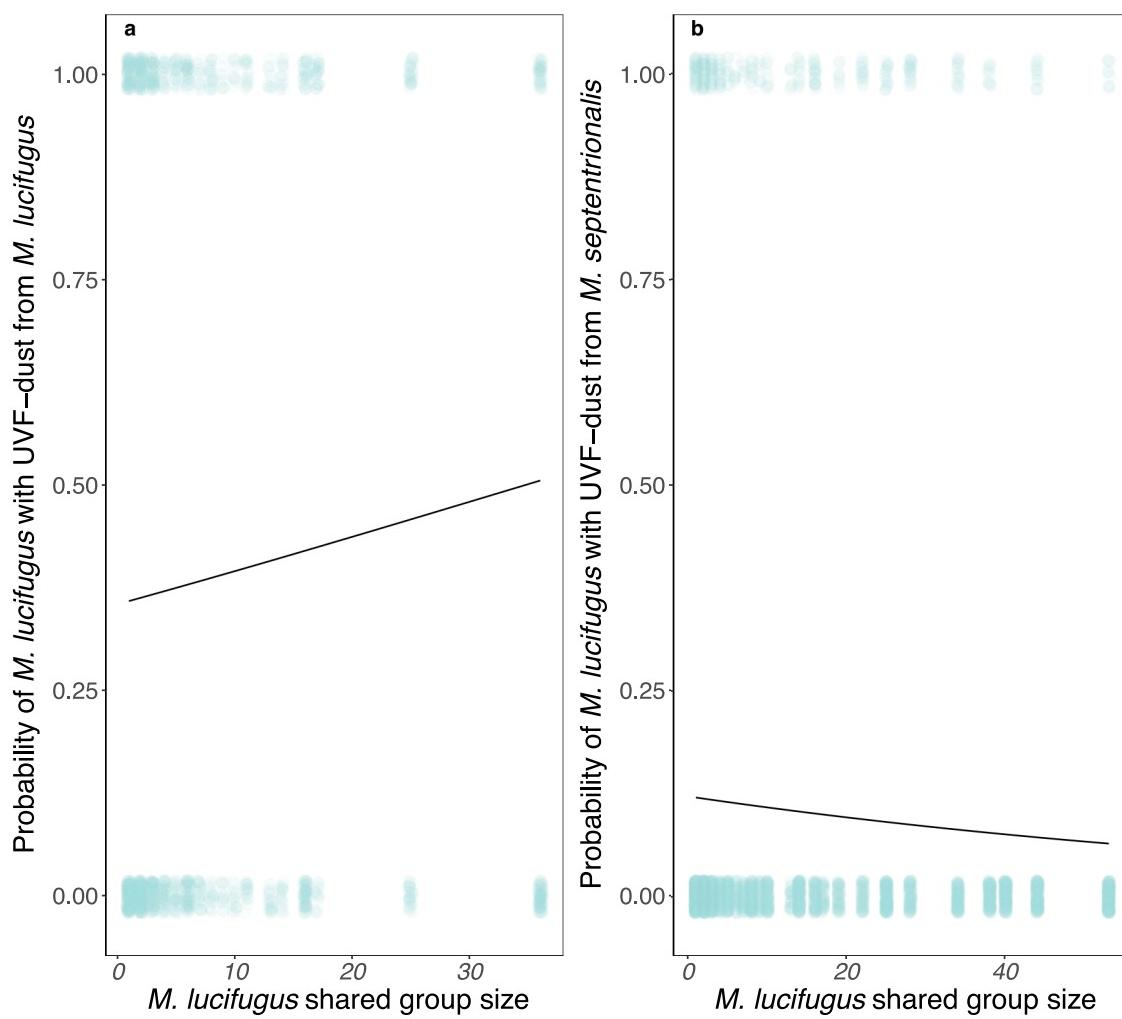
Extended Data Fig. 8 | The probability of an individual being re-sighted with UVF dust by group size. a, The probability of *M. lucifugus* (*n* = 1,697 individuals re-sighted for all dust colours) individuals having UVF dust that originated from another *M. lucifugus* increased with group size (generalized linear mixed-effects model with site and individual bat as random effects, and group size as a fixed effect; group size, coefficient: 0.017 ± 0.005; intercept, −0.596 ± 0.201, *P* = 0.0015). **b**, The probability of *M. lucifugus* (*n* = 4,164 individuals re-sighted for all dust colours) having

UVF dust that originated from dusted *M. septentrionalis* decreased with *M. lucifugus* group size (group size coefficient, −0.013 ± 0.004; intercept, −1.973 ± 0.221, *P* = 0.0003). This suggests that *M. septentrionalis* were more likely to contact *M. lucifugus* roosting solitarily than in groups, whereas *M. lucifugus* roosting in groups were more likely to be contacted by other *M. lucifugus* than were *M. lucifugus* individuals that roosted solitarily.

LETTER

Efferocytosis induces a novel SLC program to promote glucose uptake and lactate release

Sho Morioka[1,2,11], Justin S. A. Perry[1,2,11], Michael H. Raymond[1,3], Christopher B. Medina[1,2], Yunlu Zhu[4], Liyang Zhao[5], Vlad Serbulea[6], Suna Onengut-Gumuscu[7], Norbert Leitinger[6], Sarah Kucenas[4], Jeffrey C. Rathmell[8], Liza Makowski[5,9] & Kodi S. Ravichandran[1,2,10]*

Development and routine tissue homeostasis require a high turnover of apoptotic cells. These cells are removed by professional and non-professional phagocytes via efferocytosis[1]. How a phagocyte maintains its homeostasis while coordinating corpse uptake, processing ingested materials and secreting anti-inflammatory mediators is incompletely understood[1,2]. Here, using RNA sequencing to characterize the transcriptional program of phagocytes actively engulfing apoptotic cells, we identify a genetic signature involving 33 members of the solute carrier (SLC) family of membrane transport proteins, in which expression is specifically modulated during efferocytosis, but not during antibody-mediated phagocytosis. We assessed the functional relevance of these SLCs in efferocytic phagocytes and observed a robust induction of an aerobic glycolysis program, initiated by SLC2A1-mediated glucose uptake, with concurrent suppression of the oxidative phosphorylation program. The different steps of phagocytosis[2]—that is, 'smell' ('find-me' signals or sensing factors released by apoptotic cells), 'taste' (phagocyte–apoptotic cell contact) and 'ingestion' (corpse internalization)—activated distinct and overlapping sets of genes, including several SLC genes, to promote glycolysis. SLC16A1 was upregulated after corpse uptake, increasing the release of lactate, a natural by-product of aerobic glycolysis[3]. Whereas glycolysis within phagocytes contributed to actin polymerization and the continued uptake of corpses, lactate released via SLC16A1 promoted the establishment of an anti-inflammatory tissue environment. Collectively, these data reveal a SLC program that is activated during efferocytosis, identify a previously unknown reliance on aerobic glycolysis during apoptotic cell uptake and show that glycolytic by-products of efferocytosis can influence surrounding cells.

To identify the pathways potentially involved in efferocytosis, we performed RNA sequencing of LR73 hamster phagocytes engulfing apoptotic human Jurkat cells (to clearly distinguish phagocyte-derived RNA; Fig. 1a, Extended Data Fig. 1). Efferocytic phagocytes displayed changes in multiple transcriptional programs, including decreased expression of pro-inflammatory genes, increased expression of actin rearrangement/cell motility genes and increased expression of anti-inflammatory genes, consistent with previous findings[4,5] (Fig. 1a). In addition, we identified gene programs such as upregulation of glycolysis-associated genes and downregulation of genes required for oxidative phosphorylation (OXPHOS), fatty acid oxidation (FAO) and de novo cholesterol synthesis (Fig. 1a, Supplementary Table 1).

We also noted extensive modulation of genes encoding solute carrier (SLC) proteins. SLCs are membrane proteins, localized in the plasma membrane, mitochondrial membrane and other internal membranes, which facilitate transport of molecules including sugars, nucleotides and amino acids across the membrane[6–8]. Mutations in approximately 100 of the 400 known SLCs are linked to human disease[6–8]. In LR73 phagocytes, expression levels of 33 SLCs (out of 165 detected) were modified during efferocytosis: 19 were upregulated and 14 were downregulated (Extended Data Fig. 2, Supplementary Table 2).

We categorized the 33 SLCs on the basis of association with physiological processes, experimentally or by homology (Extended Data Fig. 2b), and constructed an integrated network linking each SLC and its assigned functions with other SLCs modified during efferocytosis (Fig. 1b). All but two SLCs (Slc6a4 and Slc45a4) could be clustered into eight categories: carbohydrate metabolism; intracellular pH regulation; membrane stability and volume regulation; nucleoside salvage; vitamin transport; glycosylation; amino acid transport and catabolism; and OXPHOS and FAO. This analysis also reveals coordinated regulation of SLCs linked to particular physiological function(s): multiple SLCs associated with carbohydrate metabolism were upregulated, whereas SLCs associated with OXPHOS and FAO were downregulated (Fig. 1b). The changes in SLC expression were confirmed by quantitative reverse transcription with PCR (qRT–PCR) using hamster-specific primers (Extended Data Fig. 3a, b). Thus, efferocytosis induces a specific SLC signature that potentially modulates multiple physiological processes.

Mouse peritoneal macrophages displayed similar modulation of SLCs during efferocytosis, suggesting that professional and non-professional phagocytes shared a similar response (Fig. 2a, Extended Data Fig. 3b and Supplementary Table 3). Notably, macrophages ingesting Jurkat cells coated with CD3 antibody (with comparable phagocytosis), did not show changes in the same SLCs, except for Slc29a2 and Slc29a3, which responded in the opposite direction to the efferocytosis response (Fig. 2a). Thus, in addition to the type of engulfed cell, the type of phagocytic receptor used influences the SLC program in phagocytes. Furthermore, when apoptotic Jurkat cells were injected intraperitoneally into mice, efferocytic CD11b[high]F4/80[high] macrophages exhibited similar changes in SLC gene expression to those seen in the in vitro efferocytosis (Fig. 2b).

In a time course of efferocytosis, expression of several SLCs was modulated early (0–4 h), whereas others were upregulated later (after 4 h). Even within a group of SLCs linked to a particular function, we observed early and late modulation of expression. These changes were not strict, and mRNA and protein concentrations exhibited continuous variation through the time course (Extended Data Fig. 4).

Distinct stages or phases in efferocytosis have been identified[1,2], including (i) smell phase, in which apoptotic cells and phagocytes communicate via soluble mediators; (ii) taste phase, in which ligand–receptor interactions are established between apoptotic cells and phagocytes; and (iii) ingestion phase, in which corpses are internalized and processed. To identify SLCs that are modulated during the different stages of efferocytosis, we performed and compared RNA-seq analyses

[1]The Center for Cell Clearance, University of Virginia, Charlottesville, VA, USA. [2]Department of Microbiology, Immunology and Cancer Biology, University of Virginia, Charlottesville, VA, USA. [3]Neuroscience Graduate Program, University of Virginia, Charlottesville, VA, USA. [4]Department of Biology, University of Virginia, Charlottesville, VA, USA. [5]Department of Nutrition, University of North Carolina, Chapel Hill, NC, USA. [6]Department of Pharmacology, University of Virginia, Charlottesville, VA, USA. [7]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [8]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA. [9]Department of Medicine, University of Tennessee Health Science Center, Memphis, TN, USA. [10]Inflammation Research Centre, VIB, and Department of Biomedical Molecular Biology, Ghent, Belgium. [11]These authors contributed equally: Sho Morioka, Justin S. A. Perry. *e-mail: ravi@virginia.edu
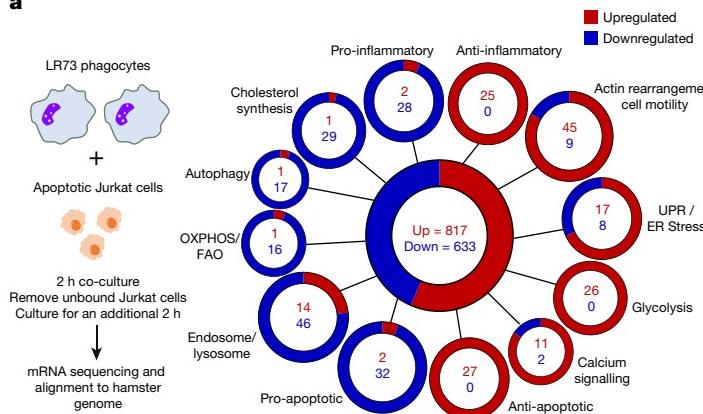
Fig. 1 | Transcriptional programs initiated during efferocytosis. a, Phagocytes regulate distinct transcriptional modules during efferocytosis. RNA-seq was performed on LR73 hamster fibroblasts that had been incubated with apoptotic human Jurkat cells. We detected changes in expression of 1,450 hamster genes. These genes were categorized according to primary function and sequence similarity.

Significance was assigned if multiple-comparisons and adjusted *P* value per DESeq2 algorithm was <0.1. Data are from four independent experimental replicates. UPR, unfolded protein response. **b**, SLC genes that are differentially regulated during efferocytosis are represented using network analysis to determine family clusters (shaded areas) and connectedness between individual SLCs.

of LR73 cells treated only with apoptotic cell supernatants with those of phagocytosing LR73 cells treated with cytochalasin D, which allows corpse binding but not internalization (Fig. 2c). These data, using SLC induction as a readout, provide further evidence for ongoing communication between apoptotic cells and phagocytes.

Efferocytosis is an energy-intensive process, as it requires energy for dynamic actin rearrangements to engulf corpses that are often nearly as large as the phagocyte[9]. We focused on SLC2A1 (also known as GLUT1), a glucose transporter that facilitates uptake of glucose from the extracellular medium[10], as SLC2A1 was strongly upregulated in LR73 cells and macrophages early during efferocytosis (Fig. 2, Extended Data Fig. 4b). First, overexpressing SLC2A1 in LR73 cells increased efferocytosis (Fig. 3a). Second, treatment with STF-31, an inhibitor of SLC2A1, reduced efferocytosis in wild-type and SLC2A1-overexpressing LR73 cells (Fig. 3a). Third, small interfering RNA (siRNA) knockdown of *Slc2a1* reduced corpse uptake (Fig. 3b); this effect was rescued by co-transfection with siRNA-resistant *Slc2a1* (Fig. 3b). Fourth, CRISPR–Cas9 deletion of *Slc2a1* reduced efferocytosis in LR73 cells (Fig. 3c). Fifth, *Slc2a1*-deficient bone-marrow-derived macrophages (BMDMs) displayed decreased efferocytosis (Fig. 3d), but retained normal antibody-mediated phagocytosis (which does not modulate *Slc2a1* expression) compared to untreated cells (Extended Data Fig. 5a). Deletion efficiencies for *Slc2a1* are shown in Extended Data Fig. 5b–d.

In vivo administration of STF-31[11] before intraperitoneal injection of apoptotic Jurkat cells decreased efferocytosis by peritoneal macrophages (Fig. 3e), but not phagocytosis of CD3-antibody-coated Jurkat cells (Extended Data Fig. 5e). Moreover, STF-31 did not further reduce efferocytosis in *Slc2a1*-deficient LR73 cells or BMDMs (Extended Data Fig. 5f), indicating that the effects were specific to SLC2A1. We also tested corpse clearance in the thymus after induction of apoptosis by dexamethasone injection[12], with or without co-injection of STF-31. There was a modest increase in uncleared or secondarily necrotic thymocytes in response to STF-31 alone, and a significant increase in necrotic thymocytes in response to treatment with both dexamethasone and STF-31 (Fig. 3f, Extended Data Fig. 5g).

To complement the pharmacological approach, we used two genetic approaches. First, using zebrafish expressing GFP in macrophages (*Tg(mpeg1:GFP)*), we targeted the *Slc2a1* orthologue via morpholino oligonucleotide injection[13]. In control morpholino-treated embryos, GFP+ macrophages exhibited numerous phagocytic puncta (containing and/or associated with neutral red), whereas *slc2a1a* morphant embryos had fewer such associations (Fig. 3g). Quantifying *z*-stack images, and focusing on macrophages in the trunk region, *slc2a1a* morphants

displayed fewer neutral red+ macrophages and less neutral-red staining per macrophage (Fig. 3g). This also suggested an evolutionarily conserved role for SLC2A1. Second, we tested the *Slc2a1* requirement in a mouse model of atherosclerosis, in which defective apoptotic cell clearance can manifest as increased necrotic cores within plaques[1,14]. Bone marrow transplantation was performed with cells from myeloid-targeted LysM–Cre *Slc2a1*^fl/fl mice (myeloid differentiation was not affected) into atherosclerosis-prone *Ldlr*−/− mice. After Western diet feeding (12 weeks), there was a significant increase of the necrotic core area in the aortic roots of mice deficient in *Slc2a1* in the myeloid lineage (Fig. 3h). The number of TUNEL-positive nuclei (late apoptotic cells) also significantly increased within necrotic cores, implicating defective corpse clearance (Fig. 3h). Taken together, these results show that SLC2A1 contributes to engulfment of apoptotic cells both in vitro and in vivo.

As SLC2A1 is a glucose transporter[8,10], we investigated whether glucose uptake is important for efferocytosis. Switching LR73 cells to glucose-free medium at the initiation of efferocytosis reduced uptake of apoptotic cells (Fig. 4a); conversely, adding exogenous glucose increased uptake of apoptotic cells (Fig. 4a), a phenotype that was attenuated by silencing *Slc2a1* (Extended Data Fig. 6a). We also measured glucose uptake directly using the non-metabolizable glucose analogue 2-deoxyglucose (2-DG); 2-DG uptake increased approximately three-fold during efferocytosis (Fig. 4b). Further, phagocytes pre-treated with 2-DG showed decreased efferocytosis (Extended Data Fig. 6b). Thus, SLC2A1-mediated glucose uptake is an important step in efferocytosis.

Seahorse analysis of LR73 phagocytes (Fig. 4c) or BMDMs (Extended Data Fig. 6c) showed increased aerobic glycolysis and decreased OXPHOS in engulfing phagocytes (Fig. 4c). Analysis of the RNA-seq data from engulfing phagocytes (Fig. 1a) revealed upregulation of multiple glycolysis genes (Fig. 4d, Extended Data Fig. 6d), with concurrent downregulation of OXPHOS and FAO genes. siRNA-mediated knockdown of PDK1 (Fig. 4e) or PDK4 (Fig. 4f), which promote aerobic glycolysis[15], or treatment with the pan-PDK inhibitor dichloroacetate (Fig. 4g), resulted in decreased efferocytosis. Notably, when we compared BMDMs treated with dichloroacetate to block aerobic glycolysis with BMDMs treated with rotenone and antimycin A1 to block OXPHOS, we found that PDK inhibition reduced efferocytosis but not antibody-mediated phagocytosis, whereas OXPHOS inhibition reduced antibody-mediated phagocytosis, but not efferocytosis (Extended Data Fig. 7a).

Phosphorylation of SLC2A1 by SGK1 increases the abundance of SLC2A1 at the plasma membrane[16]. *Sgk1* mRNA was upregulated during efferocytosis (Fig. 4d, Extended Data Fig. 6d), and targeting SGK1
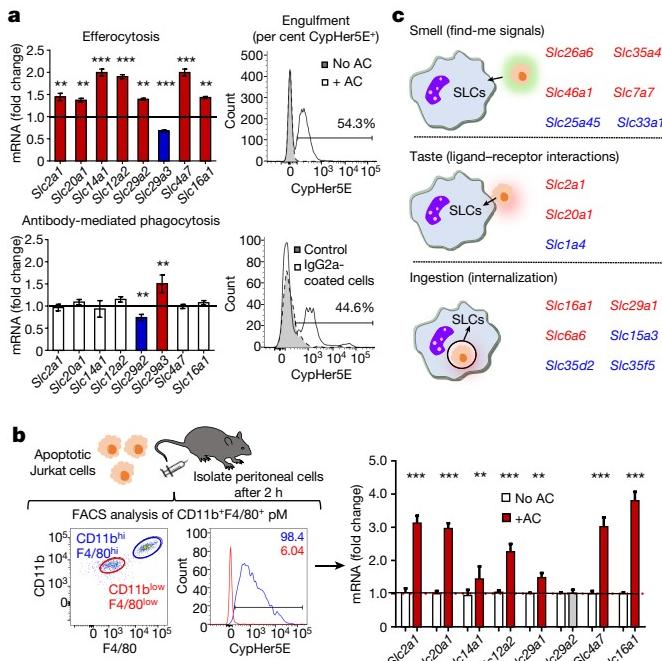
Fig. 2 | Specific SLC signatures induced during different contexts of efferocytosis. a, The SLC signature during efferocytosis is distinct from that of antibody-mediated phagocytosis. Peritoneal macrophages were incubated with apoptotic Jurkat cells (AC) or Jurkat cells coated with IgG antibody against CD3, and SLC gene expression was measured by qRT–PCR. Upregulated (green), downregulated (red) and unchanged (white) SLC expression is shown (left). CypHer5E fluorescence in macrophages engulfing targets (right). Two independent experiments with 3–4 replicates per condition. **b**, Modulation of SLC expression in efferocytic peritoneal macrophages in vivo. Flow cytometric profiles of CD11b^high^F4/80^high^ and CD11b^low^F4/80^low^ macrophages, and engulfing peritoneal macrophages (CypHer5e⁺) (bottom left). Quantification of SLC expression in the isolated CD11b^high^ fraction by qRT–PCR using mouse-specific primers (right). In **b**, right panel, the grey bar for *Slc29a2* signifies that there was no increase in expression of this SLC in the presence of apoptotic cells, relative to the control condition. Data represent two replicates with 6 mice per group per experiment. **c**, Specific SLC signature during different stages of efferocytosis. RNA-seq was performed using mRNA from LR73 cells treated for 4 h with supernatants of apoptotic cells, or LR73 cells treated with cytochalasin D and incubated with apoptotic cells. SLC genes altered by supernatant alone (smell), and cytochalasin D-sensitive SLCs (ingestion) were used to identify SLCs responding to ligand–receptor interactions (taste) (red, upregulated; blue, downregulated). For clarity, SLCs in more than one stage are not shown. In all figures, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

with siRNA or the SGK1 inhibitor GSK650394 decreased uptake of apoptotic cells (Fig. 4h, Extended Data Fig. 7b). Further, efferocytic BMDMs from transgenic mice expressing an extracellular Myc tag on SLC2A1[10] exhibited increased cell-surface expression of Myc–SLC2A1; this effect was also inhibited by GSK650394 (Extended Data Fig. 7c).

Phagocytes frequently ingest multiple apoptotic corpses sequentially[12]. To address whether SLC2A1 is required for the uptake of the first corpse, or for continued uptake of further corpses, we treated phagocytes with inhibitors for SLC2A1 or SGK1 at different times during efferocytosis. SLC2A1 function was required for uptake of the first corpse as well as for continued uptake of additional corpses (Extended Data Fig. 7d).

Corpse internalization requires substantial actin polymerization, which has been linked to aerobic glycolysis during cell migration[1,9]. Increased actin polymerization in efferocytic phagocytes (indicated by phalloidin staining) was inhibited by treatment with either STF-31 or 2-DG (Fig. 4i). Inhibiting PDK1, which favours aerobic glycolysis, also reduced F-actin formation (Fig. 4i). Thus, SLC2A1-mediated glucose
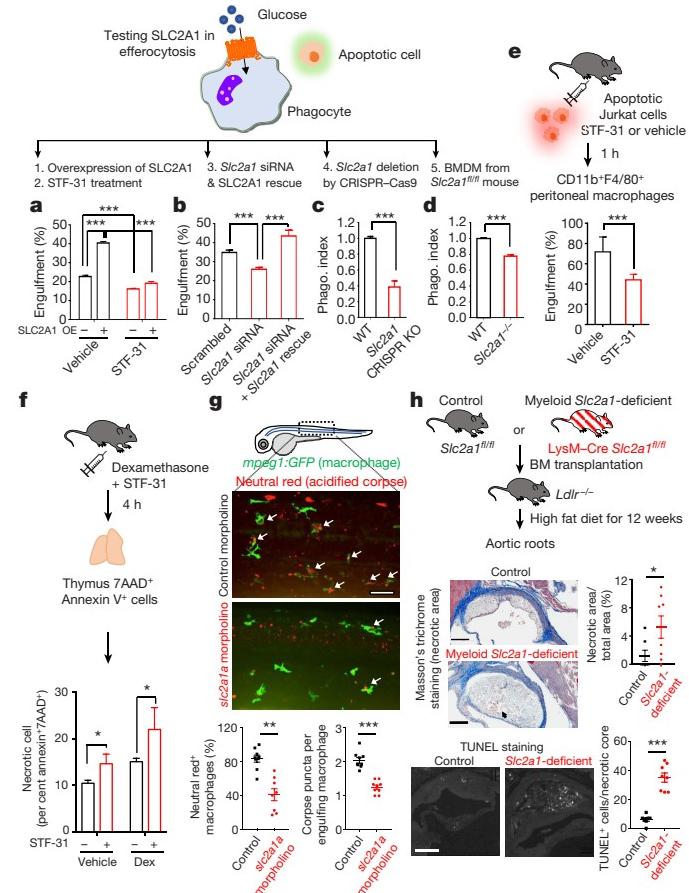
Fig. 3 | SLC2A1 promotes glucose uptake and efferocytosis. a, Overexpression of SLC2A1 in BMDMs increases efferocytosis. Treatment with the SLC2A1 inhibitor STF-31 reduces phagocytosis and abolishes the effect of SLC2A1 overexpression (OE). **b**, *Slc2a1* siRNA inhibits efferocytosis; this effect is rescued by co-transfection with siRNA-resistant *Slc2a1* cDNA. **c**, SLC2A1 knockout by CRISPR–Cas9 causes a decrease in phagocytosis of apoptotic cells in LR73 cells. WT, wild type. **d**, SLC2A1 deletion in BMDMs from *Slc2a1^fl/fl^* mice using TAT-Cre inhibits phagocytosis. Phagocytosis index = per cent engulfment (experimental/control). Data from ≥ 2 independent experiments with 3–4 replicates per condition. **e**, Treatment with the SLC2A1 inhibitor STF-31 reduces efferocytosis in vivo. **f**, STF-31 promotes accumulation of necrotic thymocytes after dexamethasone-induced apoptosis in vivo. Data represent two independent experiments with 3–4 mice per group. **g**, Targeting the *Slc2a1* orthologue reduces efferocytosis in zebrafish. *Tg(mpeg1:GFP)* embryos were injected with control or *slc2a1a* morpholino. Neutral red was used to preferentially stain acidic organelles. *slc2a1a*-targeted morphants (50 h post fertilization (hpf)) displayed less apoptotic cell engulfment (neutral red⁺ GFP-labelled macrophages) in the trunk region. Three areas and three fish per group. Data are mean ± s.d. Scale bar, 50 μm. **h**, Increased necrotic atherosclerotic area and TUNEL⁺ cells after myeloid-specific deletion of *Slc2a1*. Top, schematic of bone marrow chimaeras using *Slc2a1^fl/fl^* and LysM–Cre *Slc2a1^fl/fl^* mice. Middle, serial interrupted 5-μm sections stained with Masson's Trichrome. Representative photomicrographs and quantification of necrotic core area normalized to total area. Bottom, TUNEL staining and quantification of TUNEL⁺ cells per necrotic core. Data are mean ± s.e.m. of 7–8 mice per group. Scale bar, 200 μm.

uptake and glucose utilization in aerobic glycolysis contribute to actin polymerization during efferocytosis.

Distinct steps of SLC2A1-dependent aerobic glycolysis in phagocytes were regulated by the smell, taste and ingestion phases of efferocytosis (Fig. 5a–c). Apoptotic supernatant was sufficient to increase *Sgk1* expression, but not that of *Slc2a1* (Fig. 5a). ATP (a known find-me signal)[17] also increased *Sgk1* expression (Extended Data Fig. 8a). Similarly,
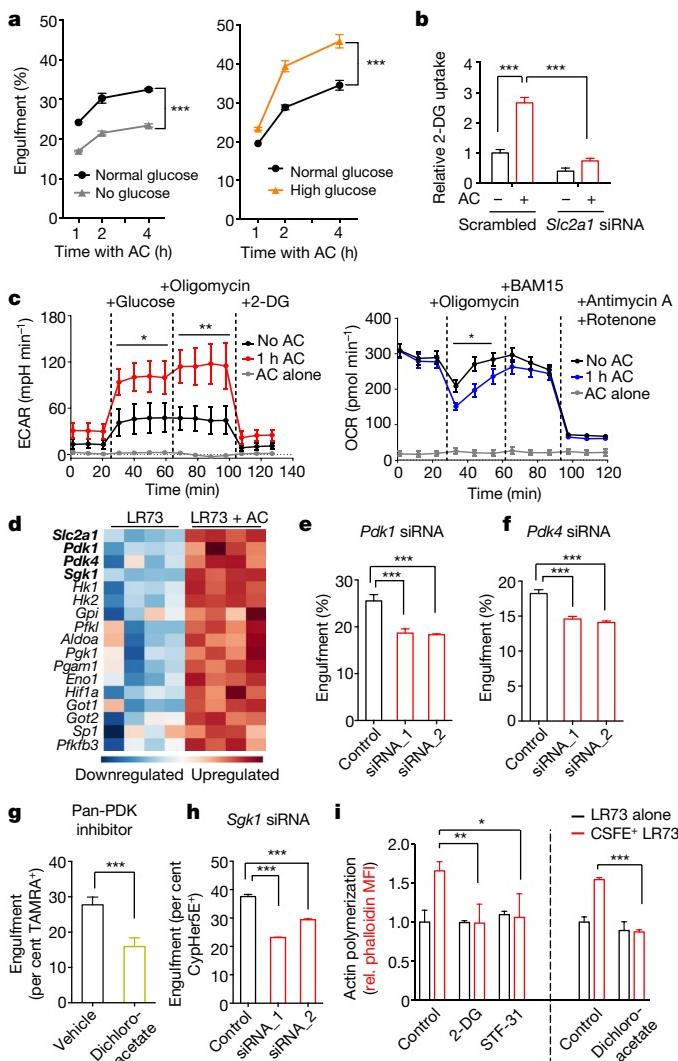
Fig. 4 | Glycolytic pathway intermediates facilitate efferocytosis.
a, Glucose in medium influences efferocytosis. No glucose, grey; physiological glucose (1 mg ml^{-1}), black; high glucose (5 mg ml^{-1}), orange. Glucose was added to medium simultaneously with apoptotic cells. Data from at least 3 independent experiments with 2–3 replicates per condition. Note that this differs from long-term glucose-free pre-treatment of phagocytes[12]. **b**, Increased glucose uptake via SLC2A1 during efferocytosis. LR73 cells were co-cultured with apoptotic Jurkat cells for 2 h, washed and 2-DG uptake was measured. Data from 2 independent experiments with 3 replicates each. **c**, Increased glycolysis in phagocytes during apoptotic cell clearance. Glycolysis and OXPHOS were measured during efferocytosis (with Seahorse XF) using extracellular acidification rate (ECAR) and oxygen consumption rate (OCR). Data are mean ± s.d. of two replicates per condition from two independent experiments. **d**, Heat map showing upregulation of glycolysis-associated genes during efferocytosis. **e–h**, Effect of siRNA targeting of *Pdk1* (**e**), *Pdk4* (**f**) or *Sgk1* (**h**), or dichloroacetate, a pan-PDK inhibitor (**g**), on efferocytosis in LR73 cells. Data represent three independent experiments with 3–4 replicates per condition. **i**, F-actin formation during efferocytosis. LR73 phagocytes were co-incubated with CFSE-labelled apoptotic thymocytes for 30 min, stained with phalloidin and actin polymerization was measured using by flow cytometry. PDK1 inhibitor experiments were performed separately. MFI, mean fluorescence intensity; rel., relative. Data represent at least 3 independent experiments with 3–4 replicates.

binding of apoptotic targets or phosphatidylserine (PtdSer) liposomes to phagocytes (without internalization) was sufficient to induce *Slc2a1* (Fig. 5b). Masking PtdSer on targets (using BAI1-TSR, the PtdSer-binding domains of BAI1) reduced upregulation of *Slc2a1* (but not that of *Sgk1*) during efferocytosis (Extended Data Fig. 8b). Therefore,
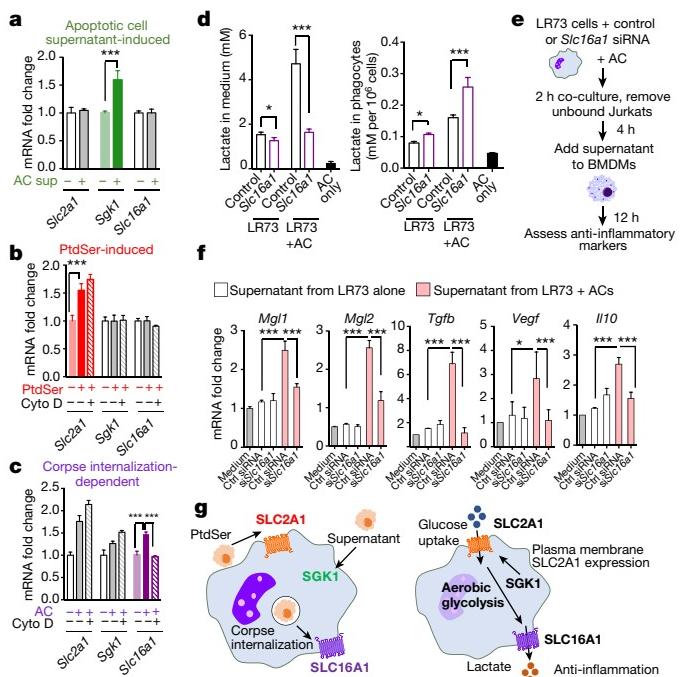
Fig. 5 | SLC16A1-mediated lactate release promotes an anti-inflammatory environment. a–c, Expression of *Slc2a1*, *Slc16a1* and *Sgk1* during distinct steps of efferocytosis, determined by qRT–PCR. mRNA induction by supernatant of apoptotic cells (**a**), PtdSer-containing liposomes (**b**) or apoptotic cells with or without cytochalasin D (1 μM) for 4 h (**c**), which were added to LR73 cell cultures. **d**, Lactate release from efferocytic phagocytes. LR73 cells treated with control or *Slc16a1* siRNA were incubated with apoptotic cells, washed, incubated for an additional 4 h, and lactate was measured in the cells and in the medium. Data in **a**–**d** represent 2 independent experiments with 3–4 replicates per condition. **e, f**, Supernatant from LR73 phagocytes (**e**) was added to BMDMs, and the effect on expression of anti-inflammatory or M2-like marker genes was analysed by qRT–PCR (**f**). Data represent 2 independent experiments with 2–3 replicates per condition. **g**, Schematic showing the upregulation and function of the SLC2A1–SGK1–SLC16A1 axis during efferocytosis.

SGK1—triggered by factors released from apoptotic cells—prepares the phagocytes by increasing abundance of endogenous SLC2A1 on the plasma membrane, and PtdSer-dependent interactions increase new transcription and expression of SLC2A1.

Following corpse internalization in phagocytes, there was increased expression of *Slc16a1*, a plasma membrane proton-linked monocarboxylate transporter of lactate and pyruvate[18] (Figs. 1b, 2a–c, 5c, Extended Data Fig. 8b). siRNA-mediated knockdown of *Slc16a1* in LR73 cells reduced efferocytosis in vitro (Extended Data Fig. 8c), and SR13800—a bioactive inhibitor of SLC16A1—reduced apoptotic cell uptake by peritoneal macrophages in vivo (Extended Data Fig. 8d). Supernatants of engulfing LR73 cells contained threefold-higher lactate concentration (approximately 5 mM) compared to supernatants of phagocytes without apoptotic cells (approximately 1.5 mM) (Fig. 5d). siRNA knockdown of *Slc16a1* reduced lactate concentration in the supernatants to approximately 2 mM, with concomitant accumulation in phagocytes (Fig. 5d), indicating that SLC16A1 contributes to lactate release from engulfing phagocytes.

Lactate released from solid tumours could act on naive macrophages to induce M2 macrophage-like polarization[3]. To test whether factors released via SLC16A1 during efferocytosis might also promote anti-inflammatory skewing of naive macrophages[19], we tested the effect of supernatants of engulfing LR73 phagocytes on BMDMs (Fig. 5e). These supernatants induced upregulation of anti-inflammatory macrophage genes such as *Tgfb1* and *Il10* as well as anti-inflammatory or M2-like markers, including *Vegfa*, *Mgl1*, *Mgl2* and *CD206* (also known as *Mrc1*),

whereas pro-inflammatory markers (*Tnf* and *Il6*) were not affected. *Slc16a1* knockdown attenuated this effect in engulfing LR73 cells (Fig. 5f, Extended Data Fig. 8e, f). Therefore, aerobic glycolysis induced during efferocytosis affects efferocytosis at two levels: regulation of corpse uptake through regulation of actin polymerization (involving SLC2A1); and modulation of expression of anti-inflammatory genes in neighbouring cells (via SLC16A1).

The results presented here provide several key insights. Although the SLC family is the second largest among membrane proteins (after G-protein-coupled receptors), there is much less knowledge of SLC function in specific physiological contexts[6]. This work describes coordinated regulation of select SLCs during efferocytosis (Fig. 5g). The smell, touch and ingestion phases of efferocytosis induce distinct and overlapping sets of SLC genes with functional consequences; these sets of genes are distinct from those that are activated during antibody-mediated phagocytosis. Efferocytosis induces a metabolic gene program promoting glucose uptake and subsequent glycolysis, with concurrent downregulation of genes linked to OXPHOS and FAO. Although efferocytic macrophages are more M2-like[20], and M2-like macrophages are reported to be OXPHOS-dependent[21], our study of the first few hours of efferocytosis does not rule out a role for OXPHOS at later times. Although glycolysis is linked to inflammation, efferocytic phagocytes can influence non-engulfing naive macrophages in the tissue microenvironment towards anti-inflammatory polarization by SLC16A1-mediated lactate release, as well as other factors[22] released during efferocytosis such as TGFβ and IL-10.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0735-5.

1. Arandjelovic, S. & Ravichandran, K. S. Phagocytosis of apoptotic cells in homeostasis. *Nat. Immunol.* **16**, 907–917 (2015).
2. Elliott, M. R. & Ravichandran, K. S. The dynamics of apoptotic cell clearance. *Dev. Cell* **38**, 147–160 (2016).
3. Colegio, O. R. et al. Functional polarization of tumour-associated macrophages by tumour-derived lactic acid. *Nature* **513**, 559–563 (2014).
4. A-Gonzalez, N. et al. Phagocytosis imprints heterogeneity in tissue-resident macrophages. *J. Exp. Med.* **214**, 1281–1296 (2017).
5. Cummings, R. J. et al. Different tissue phagocytes sample apoptotic cells to direct distinct homeostasis programs. *Nature* **539**, 565–569 (2016).
6. Cesar-Razquin, A. et al. A call for systematic research on solute carriers. *Cell* **162**, 478–487 (2015).
7. Lin, L., Yee, S. W., Kim, R. B. & Giacomini, K. M. SLC transporters as therapeutic targets: emerging opportunities. *Nat. Rev. Drug Discov.* **14**, 543–560 (2015).
8. Mueckler, M. & Thorens, B. The SLC2 (GLUT) family of membrane transporters. *Mol. Aspects Med.* **34**, 121–138 (2013).
9. Verdone, J. E., Zarif, J. C. & Pienta, K. J. Aerobic glycolysis, motility, and cytoskeletal remodeling. *Cell Cycle* **14**, 169–170 (2015).
10. Macintyre, A. N. et al. The glucose transporter Glut1 is selectively essential for CD4 T cell activation and effector function. *Cell Metab.* **20**, 61–72 (2014).
11. Chan, D. A. et al. Targeting GLUT1 and the Warburg effect in renal cell carcinoma by chemical synthetic lethality. *Sci. Transl. Med.* **3**, 94ra70 (2011).
12. Park, D. et al. Continued clearance of apoptotic cells critically depends on the phagocyte Ucp2 protein. *Nature* **477**, 220–224 (2011).
13. Jensen, P. J., Gitlin, J. D. & Carayannopoulos, M. O. GLUT1 deficiency links nutrient availability and apoptosis during embryonic development. *J. Biol. Chem.* **281**, 13382–13387 (2006).
14. Kojima, Y., Weissman, I. L. & Leeper, N. J. The role of efferocytosis in atherosclerosis. *Circulation* **135**, 476–489 (2017).
15. Zhang, S., Hulver, M. W., McMillan, R. P., Cline, M. A. & Gilbert, E. R. The pivotal role of pyruvate dehydrogenase kinases in metabolic flexibility. *Nutr. Metab.* **11**, 10 (2014).
16. Palmada, M. et al. SGK1 kinase upregulates GLUT1 activity and plasma membrane expression. *Diabetes* **55**, 421–427 (2006).
17. Elliott, M. R. et al. Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance. *Nature* **461**, 282–286 (2009).
18. Sprowl-Tanio, S. et al. Lactate/pyruvate transporter MCT-1 is a direct Wnt target that confers sensitivity to 3-bromopyruvate in colon cancer. *Cancer Metab.* **4**, 20 (2016).
19. Biswas, S. K. & Mantovani, A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. *Nat. Immunol.* **11**, 889–896 (2010).
20. Kelly, B. & O'Neill, L. A. Metabolic reprogramming in macrophages and dendritic cells in innate immunity. *Cell Res.* **25**, 771–784 (2015).
21. Gaber, T., Strehl, C. & Buttgereit, F. Metabolic regulation of inflammation. *Nat. Rev. Rheumatol.* **13**, 267–279 (2017).
22. Han, C. Z. et al. Macrophages redirect phagocytosis by non-professional phagocytes and influence inflammation. *Nature* **539**, 570–574 (2016).

Author contributions S.M. and J.S.A.P. designed and performed most experiments, with input from K.S.R. M.H.R., C.B.M., V.S., N.L., S.O.-G., J.C.R., Y.Z., S.K., L.Z. and L.M. performed and/or assisted with specific experiments. S.M., J.S.A.P. and K.S.R. wrote the manuscript with input from co-authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0735-5.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0735-5.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to K.S.R.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

The investigators were not blinded to allocation during experiments and outcome assessment.

In vitro engulfment assay. For induction of apoptosis[22], human Jurkat T cells resuspended in RPMI with 1% BSA were treated with 150 mJ/cm^2 ultraviolet C irradiation (Stratalinker) and incubated for 4 h at 37 °C with 5% CO_2. For antibody-dependent phagocytosis, Jurkat cells were labelled with CD3 antibody (25 μg/ml; clone OKT3, BioLegend) along with annexin V recombinant protein (to block efferocytosis of any residual dying cells) (3 μg/ml; eBioscience) for 1 h at 4 °C and the cells[23] were then stained with CypHer5E (GE Healthcare, PA15401) or TAMRA (Invitrogen, C-1171) before use in the engulfment assays. Chinese hamster LR73 cells or mouse macrophages were seeded in a 24-well plate and incubated with targets at a 1:10 phagocyte:target ratio for the indicated times. Targets were then washed with PBS. Where indicated, phagocytes were rested in culture medium for an additional period of time. Cells were dissociated from the plate with trypsin and the phagocytes were assessed by a flow cytometry-based assay or analysed for RNA or protein[24]. Phalloidin staining was conducted according to the manufacturer's instruction (Invitrogen). When primary macrophages were used as phagocytes, there was an inherent difference in the absolute percentage uptake of corpses between experiments performed on different days. Therefore, phagocytic index was used to better compile data from multiple experiments. All cell lines (except LR73 cells) were obtained from ATCC. LR73 cells were obtained from C. Stanners[23]. Cells we routinely tested for mycoplasma contamination and all cell lines tested negative.

RNA sequencing. LR73 cells were co-cultured with apoptotic Jurkat cells for 2 h, unbound Jurkat cells were removed by washing with PBS and the phagocytes were rested in culture medium for an additional 2 h. Total RNA was extracted, and an mRNA library was prepared using the Illumina TruSeq platform, followed by sequencing using an Illumina NextSeq 500 cartridge. Four independent experiments were sequenced. R v.3.2.2 was used for graphical and statistical analysis and the R package DESeq2 was used for count normalization and differential gene expression analysis of RNA-seq data. All genes were curated using a combination of literature mining and function determination (known or predicted) via Uniprot. Genes involved in multiple associated functions (for example, kinases or cell-cycle-related) were excluded. We generated functional classifications of solute carrier (SLC) proteins on the basis of several criteria: (1) a known function was ascribed to the channel directly (for example, SLC2A1 is required for glycolysis); a known function was ascribed to the solute transported by the channel (for example, SLC2A1 transports glucose, glucose is required for glycolysis, therefore SLC2A1 is required for glycolysis); or by predictive homology (for example, SLC2A1 transports glucose and is required for glycolysis; as SLC2A4 is homologous to SLC2A1, SLC2A4 may therefore be important for glycolysis). Several SLCs have been shown or are predicted to perform multiple functions. These alternative functions are listed in Supplementary Table 3. Network analysis was performed using the network analytical software Gephi v.0.9.1 (https://gephi.org/). Standard algorithms for calculating clusters were used as implemented in Gephi. The edges and community links were calculated using the 'link communities' plug-in based on the algorithms proposed for biological network analyses[25]. Network structure was determined using the Fruchterman–Reingold force-directed layout algorithm. All code used for analysis is available upon request.

qRT–PCR. Total RNA was extracted from cells using the RNeasy Mini Kit (Qiagen) and cDNA was synthesized using QuantiTect Reverse Transcription Kit (Qiagen), according to the manufacturers' instructions. Quantitative gene expression analysis for hamster and mouse SLCs was performed using hamster- and mouse-sequence specific Taqman probes that are non-cross reactive with human sequences (Applied Biosystems), run on a StepOnePlus Real Time PCR System (Applied Biosystems). Details of Taqman primers are listed in Supplementary Information.

In vivo engulfment assay and qRT–PCR. Six million apoptotic Jurkat cells, stained with CypHer5E, or as control X-VIVO 10 medium alone, were intraperitoneally injected (300 μl per mouse). At indicated times post-injection, mice were euthanized, and the peritoneal lavage was collected in 10 ml PBS + 10% fetal bovine serum (FBS). The collected cells were stained with CD11b PE-Cy7 (eBioscience, Cat#: 25-0112-82) and F4/80 APC-eFluor 780 (eBioscience, Cat#: 47-4801-80), and the uptake of the injected CypHer5E$^+$ apoptotic cells by CD11b$^+$F4/80^{high} cells was assessed by flow cytometry. For sequencing or other types of analysis of the responses of macrophages, the peritoneal macrophages were isolated using Macrophage Isolation Kit (Miltenyi Biotec). To test the effect of drugs targeting SLC2A1 and SLC16A1 on in vivo engulfment, mice were treated with STF-31 (Santa Cruz #sc-364692, 10mg/kg) or SR13800 (EMD Millipore # 509663, 10mg/kg) for 1 h before administration of apoptotic cells or IgG-coated cells. For qRT–PCR, the cells collected were lysed for RNA isolation.

In vivo thymus efferocytosis assay. Six- to eight-week-old mice were injected intraperitoneally with 300 μl PBS containing 250 μg dexamethasone (Sigma) with or without STF-31 (10 mg/kg) dissolved in EtOH. Four hours after injection, thymi were collected from mice, and the numbers of thymocytes with annexin V staining only (apoptotic) versus annexin V, 7-aminoactinomycin D (7-AAD) double-positive cells (secondarily necrotic) were determined by flow cytometry.

In vivo bone marrow transplantation and analysis of necrotic cores in atherosclerosis. At six weeks of age, $Ldlr^{-/-}$ mice received two doses of X-ray irradiation (500 cGy × 2, 4 h apart; X-RAD), and were then transplanted with bone marrows isolated from $Slc2a1^{fl/fl}$ or $LysM$-cre/$Slc2a1^{fl/fl}$ donor mice. Control animals were transplanted with HBSS buffer only—these mice died within 10–11 days of lethal irradiation. Following bone marrow transplantation, chimeric $Ldlr^{-/-}$ mice were transferred to sterile cages with ad libitum access to sterile mouse chow and sterile water and were maintained on chow diet for four weeks before challenge with Western Diet (Harlan Teklad TD88137, 42% of Kcal from milk fat with 0.15% added cholesterol) for 12 weeks. Following mice dissection, hearts were isolated for formalin-fixed paraffin-embedded sections. The sections were stained with Masson's trichrome for quantification of necrotic core areas and normalized to total quantified area. Necrotic core size was measured using Aperio ePathology software (Aperio). TUNEL staining was conducted according to the manufacturer's instructions (Promega).

In vivo phagocytosis analysis in zebrafish. To inhibit the expression of slc2a1, antisense morpholinos targeting the translational start site of the zebrafish $slc2a1a$ (5′–GGCCATCATCAGCTGAGGAGTCACC–3′) were synthesized (Gene Tools). A control morpholino (5′ GGCCATCATCAGCTGAGGAGTCACC–3′) was used as a negative control. Morpholino (2.5 ng) was microinjected into $Tg(mpeg1$:$GFP)$ embryos at the one-cell stage. Embryos were treated with phenylthiourea (0.004%) in egg water at 24 hpf to reduce pigmentation as per standard protocols. 8–10 h before imaging, embryos were soaked in 2.5 μg/ml neutral red (Sigma) in egg water. At 50 hpf, morphants were anaesthetized with 3-aminobenzoic acid ester (Tricaine), immersed in 0.8% low-melting-point agarose and mounted laterally in glass-bottomed 35-mm Petri dishes for confocal imaging. Three identical z-stack images were taken for each embryo covering hemi-segments of somites 6 to 20. More than 100 macrophages were counted per group for the evaluation of GFP and neutral red colocalization.

CRISPR–Cas9 deletion or siRNA knockdown of SLCs. Stable, individual clones of Cas9–GFP-expressing LR73 cells were generated by lentiCas9–EGFP plasmid via lentiviral transduction and a protocol adapted from ref. [26], followed by single-cell cloning of GFP-expressing cells and Cas9 expression verification. $Slc2a1$ was deleted from LR73 cells using two independent Cas9-GFP LR73 cell clones and using lentiGuide-Puro sgRNA plasmid with two unique guides for $Slc2a1$. LentiCas9–EGFP was a gift from P. Sharp and F. Zhang (Addgene plasmid # 63592) and lentiGuide-Puro was a gift from F. Zhang (Addgene plasmid # 52963).

Guide RNAs targeting $Slc2a1$ were generated using the following oligonucleotide pairs. Guide 1: 5′ CACCGATTCTTCCGGACATCATCGC–3′, 3′–CTAA GAAGGCCTGTAGTAGCGCAAA–5′; guide 2: 5′–CACCGTTCGGCC TGGACTCCATTA–3′, 3′–CAAGCCGGACCTGAGGTAATCAAA–5′.

For siRNA and plasmid transduction experiments, LR73 cells were treated with Lipofectamine 2000 (Thermo Fisher) with specific siRNAs, according to the manufacturer's instructions, 2 days before the engulfment assay. GLUT1–eGFP/pcDNA-DEST47 was a gift from W. Frommer (Addgene plasmid #18729)[27]. siRNAs targeting hamster mRNAs were customized by GE Healthcare Dharmacon.

Primers for siRNA against $Slc2a1$: 5′–CCAAGAGUGUGCUGAAGAAUU–3′. Two siRNAs against $Sgk1$: 5′–CUUCUAUGCUGCUGAAAUAUU–3′, 5′–CU GCAGAAGGACAGGACAAUU–3′. Two siRNAs against $Pdk1$: 5′–CGA CAAGAGUUGCCUGUUAUU–3′, 5′–GGACAAAAGUGCUGAAGAUUU–3′. Two siRNAs against $Pdk4$: 5′–UCACACAAGUAAAUGGAAAUU–3′, 5′–CAUCAAAGUUCGAAACAGAUU–3′. siRNA against $Slc16a1$: 5′–AGAAA CAGGAAGAAGGUAAUU–3′.

Macrophage isolation and analysis. To obtain BMDMs, femurs from control mice, mice carrying floxed alleles of $Slc2a1$, or Glut1–Myc knock-in mice were removed and flushed with 5 ml sterile PBS containing 5% FBS[28,29]. The cell suspension was centrifuged, treated with red blood cell lysis buffer, washed and then plated onto sterile Petri dishes in DMEM containing 10% L929 medium, 10% FBS and 1% penicillin–streptomycin–glutamine. Medium was replenished every 2–3 days and differentiated cells were used at day 6 post-collection. To delete $Slc2a1$, macrophage cultures were treated with TAT-Cre (EMD Millipore), according to the manufacturer's instructions. For staining, BMDMs were stained with F4/80 APC-eFluor 780 (eBioscience, Cat#: 47-4801-80) and subsequently stained with Myc PerCP antibody (Novus Biologicals, 9E10 Cat#: NB600-302PCP) or fixed and permeabilized using FoxP3/Transcription Factor Staining Buffer Set (eBioscience), and intracellular staining was performed using CD206 PE (BioLegend, Cat#: 141706). Resident peritoneal macrophages were obtained by flushing the peritoneal cavity of mice with 10 ml cold PBS containing 5% FBS. Collected cells were spun down, resuspended in X-VIVO 10 (Lonza), and plated at a concentration of 5×10^5 cells per well. Floating cells were removed the next day, and remaining peritoneal macrophages were used 2 days after isolation.

Glucose uptake assay. LR73 cells were incubated with apoptotic Jurkat cells for 2 h, washed 3 times with PBS and incubated with 10 mM 2-DG, a glucose analogue, in glucose-free medium for 30 min. Following incubation, cells were washed 3 times with PBS and lysed with Extraction Buffer (Sigma Cat#: MAK083). Lysate was frozen–thawed in dry ice/ethanol, and then heated at 85 °C for 40 min. Lysate was then cooled on ice for 5 min and then neutralized by Neutralization Buffer (Sigma Cat#: MAK083). Samples were spun down at 13,000g to remove insoluble fraction and then diluted tenfold by adding Assay Buffer (Sigma Cat#: MAK083). Using the lysate, glucose uptake was measured using Glucose Uptake Colorimetric Assay Kit (Sigma). 2-DG is taken up by cells and phosphorylated by hexokinase to 2-DG6P. 2-DG6P cannot be further metabolized and accumulates in cells, directly proportional to the glucose uptake by cells. 2-DG uptake is determined by a coupled enzymatic assay in which the 2-DG6P is oxidized, resulting in the generation of NADPH, which is then determined by a recycling amplification reaction in which the NADPH is used by glutathione reductase in a coupled enzymatic reaction that produces glutathione. Glutathione reacts with DTNB to produce TNB, which was detected at 412 nm as per the manufacturer's recommendations.

Seahorse analysis. LR73 cells or BMDMs were seeded into a Seahorse 24-well tissue culture plate (Agilent Technologies). The cells adhered overnight before treatment. For assessing respiratory capacity, cells were subjected to a mitochondrial stress test. In brief, at the beginning of the assay, the medium was changed to DMEM containing pyruvate (Thermo Fisher Cat#:12800017, pH = 7.35 at 37 °C) and cells were allowed to equilibrate for 30 min. OCR was measured using a Seahorse XF24 Flux Analyzer (Agilent Technologies). After three basal OCR measurements, the drugs of interest were injected into the plate and OCR was measured using four-minute measurement periods. Compounds to modulate cellular respiratory function (1 μM oligomycin (Sigma-Aldrich); 2 μM BAM15 (Cayman Chemical Company); 1 μM antimycin A and 100 nM rotenone (Sigma-Aldrich)) were injected after every three measurements. Basal respiration was calculated by subtracting the average of the first three measurements by the average of the post-antimycin A and rotenone measurements. Maximum respiratory capacity was calculated by subtracting the average of the post-BAM15 measurements by the average of the post- antimycin A and rotenone measurements. The reserve capacity was calculated by subtracting the average of the basal measurements from the average of the post-BAM15 measurements.

For assessing glycolytic capacity, the cells were subjected to a glycolytic stress test. In brief, ECAR—a measurement of lactate export—was measured using a Seahorse XF24 Flux Analyzer. Cells were seeded into a Seahorse 24-well tissue culture plate. At the beginning of the assay, the medium was changed to unbuffered, glucose-free DMEM (Sigma-Aldrich Cat# D5030, pH 7.35 at 37 °C), supplemented with 143 mM NaCl and 2 mM glutamine. After three basal ECAR measurements, the drugs of interest were injected into the plate and ECAR was measured using 3-min measurement periods. Compounds to modulate glycolysis (20 mM glucose; 1 μM oligomycin; 80 mM 2-DG) (Sigma) were injected after every three measurements. Basal glycolysis was calculated by subtracting the average of the post-2-DG measurements from the average of the post-glucose measurements. Maximum glycolytic capacity was calculated by subtracting the average of the post-2-DG measurements from the average of the post-oligomycin measurements. The glycolytic reserve capacity was calculated by subtracting the average of the post-oligomycin measurements from the average of the post-glucose measurements.

Liposome construction. Liposomes were prepared by dissolving the lipids (phosphatidylserine, dioleoyl phosphatidylcholine, cholesterol and the lipid DiD dye) in chloroform, evaporating chloroform under flow of argon gas in a glass vial and subjecting the lipid layer to overnight lyophilization to remove traces of organic solvent. Normal saline was then added for hydration, and after vortexing was done to prepare multilamellar vesicles. Particle size was verified by dynamic light scattering using Nicomp 370.

Determination of lactate concentration. The lactate concentration was measured using a Lactate Assay Kit (Sigma) according to the manufacturer's instructions. The mean values ± s.d. of the lactate concentration in the medium and cells were calculated for each condition.

Immunoblotting. LR73 cells were seeded in a 100-mm dish at a concentration of 2 million cells per dish. Apoptotic Jurkat cells were added as indicated. Cells were lysed in RIPA buffer and immunoblotted with SLC2A1 (Abcam #ab652), SLC16A1 (LSBio LS-C335287), SLC12A2 (Cell Signaling Technology #14581) and total Erk2 (Santa Cruz Biotechnology, #sc-154-G) antibodies in Can Get Signal solution (TOYOBO Cat# NKB-101) followed by chemiluminescence detection. Specific bands were quantified using Adobe Photoshop CS6.

Research animals. Power of 80% was used to determine the number of mice needed to achieve a two-sided 5% significance level to detect a twofold change for each set of in vivo studies. Allocation of mice was random in all in vivo experiments. Mice were taken from littermates. Animal breeding and experiments were performed in a specific pathogen-free animal facility using protocols approved by th University of Virginia Animal Studies Committee. Ethical guidelines determined by the Institutional Animal Care and Use Committee were followed in all experiments performed in this manuscript.

Statistical analysis. Statistical significance was determined using GraphPad Prism 7, using unpaired Student's two-tailed t-test, one-way ANOVA or two-way ANOVA, according to test requirements. Grubbs' outlier test was used to determine outliers, which were excluded from final analysis. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ were considered significant. Number of replicates and repeats of individual experiments and statistical tests used are shown in Supplementary Table 4.

Code availability. All code used in this manuscript can be accessed from the Github repository https://github.com/perryjs/Perry-R.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

RNA sequencing data presented in this study have been deposited in the NCBI GEO repository under the accession number GSE119273.

23. Pollard, J.W. & Stanners, C.P. Characterization of cell lines showing growth control isolated from both the wild type and a leucyl-tRNA synthetase mutant of Chinese hamster ovary cells. *J. Cell. Physiol.* **98**, 571–585 (1979).
24. Fond, A. M., Lee, C. S., Schulman, I. G., Kiss, R. S. & Ravichandran, K. S. Apoptotic cells trigger a membrane-initiated pathway to increase ABCA1. *J. Clin. Invest.* **125**, 2748–2758 (2015).
25. Ahn, Y. Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature.* **466**, 761–764 (2010).
26. Chen, S. et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell.* **160**, 124–1260 (2015).
27. Takanaga, H., Chaudhuri, B. & Frommer, W. B. GLUT1 and GLUT9 as major contributors to glucose influx in HepG2 cells identified by a high sensitivity intramolecular FRET glucose sensor. *Biochim. Biophys. Acta.* **1778**, 1091–1099 (2008).
28. Macintyre, A. N. et al. The glucose transporter Glut1 is selectively essential for CD4 T cell activation and effector function. *Cell Metab.* **20**, 61–72 (2014).
29. Michalek, R. D. et al. Cutting edge: distinct glycolytic and lipid oxidative metabolic programs are essential for effector and regulatory CD4+ T cell subsets. *J. Immunol.* **186**, 3299–3303 (2011).

a

**Phagocyte samples used
RNAseq using hamster LR73**

Apoptotic Jurkat cells

b

Quality control statistic for RNAseq
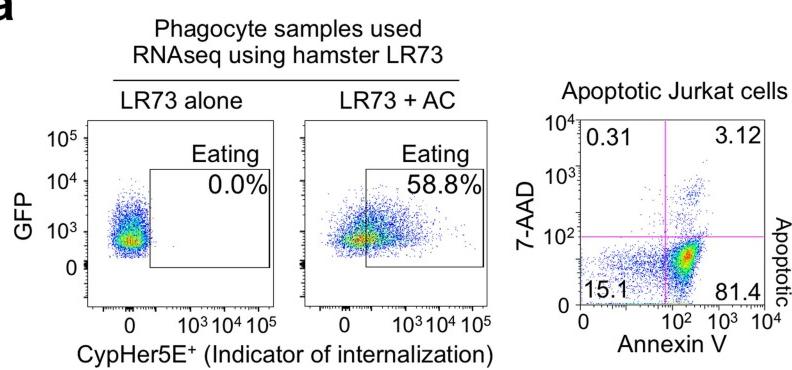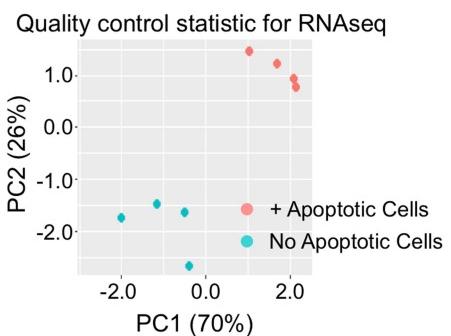
Extended Data Fig. 1 | RNA preparation for RNA-seq experiments.
a, Representative fluorescence-activated cell sorting plots of engulfment
assays with LR73 hamster phagocytes (left) and annexin V/7-AAD
staining of apoptotic human Jurkat cells (right) in conditions matching
experiments performed for RNA-seq (2 h with apoptotic cells followed by
2 h rest in the absence of apoptotic cells). **b**, Principal component analysis
on hamster-genome-aligned RNA-seq data as a quality control statistic.

a

Solute carrier (SLC) genes detected via RNAseq of engulfing cells (165 SLCs in LR73)

Upregulated during engulfment (19 genes)

Log(adj. p value)

Alphanumeric SLC identity

Down-regulated during engulfment (14 genes)
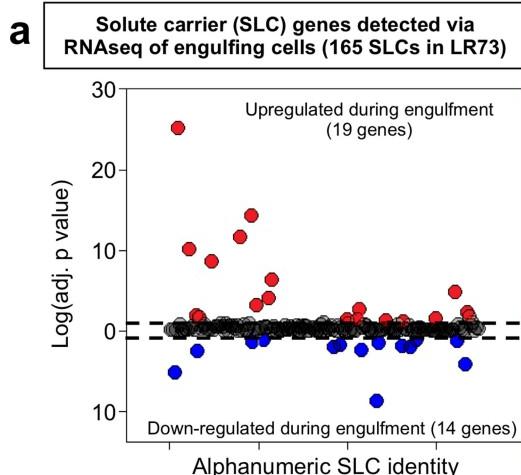
SLC	Gene family determination	Adj. p value
2a1	Facilitative GLUT Transporter	8.07E-26
4a7	Bicarbonate Transporter	7.95E-11
5a6	Sodium Glucose Cotransporter	1.30E-02
6a6	Na & Cl-dependent neurotransmitter transporter	2.66E-02
7a7	Cationic AA Transporter/Glycoprotein-associated	2.97E-09
12a2	Electroneutral cation-coupled Cl cotransporter	2.61E-12
14a1	Urea Transporter	5.21E-15
16a1	Monocarboxylate Transporter	7.27E-04
19a1	Folate/Thiamine Transporter	1.05E-04
20a1	Phosphate Transporter	5.63E-07
26a6	Multifunctional Anion Exchanger	4.45E-02
29a1	Facilitative-Nucleoside Transporter	3.67E-02
29a2	Facilitative-Nucleoside Transporter	2.24E-03
35a4	Nucleoside-Sugar Transporter	5.56E-02
35e1	Nucleoside-Sugar Transporter	7.60E-02
39a10	Metal Ion Transporter	2.84E-02
41a2	MgtE-like Magnesium Transporter	1.53E-05
45a4	Putative Sugar Transporter	5.56E-03
46a1	Folate Transporter	1.93E-02

SLCs (n=132) unchanged during apoptotic cell clearance

SLC	Gene family determination	Adj. p value
1a4	High-affinity glutamate and neutral AA Transporter	6.46E-06
6a4	Na & Cl-dependent neurotransmitter Transporter	2.65E-03
15a3	Proton Oligopeptide Cotransporter	3.67E-02
36a1	Proton-coupled AA Transporter	6.94E-02
33a1	Acetyl-CoA Transporter	3.46E-02
45a1	Putative Sugar Transporter	6.53E-05
17a5	Vesicular Glutamate Transporter	7.38E-02
25a42	Mitochondrial Carrier	8.52E-03
25a45	Mitochondrial Carrier	1.67E-02
29a3	Nucleoside-Sugar Transporter	3.52E-03
31a2	Copper Transporter	1.73E-09
35d2	Nucleoside-Sugar Transporter	1.39E-02
35f5	Nucleoside-Sugar Transporter	1.05E-02
41a3	MgtE-like Magnesium Transporter	5.76E-02

b SLC categorization

| Gene	Primary Function	SLC Family	Monogenic Disease?	SNP/Phenotype?
Slc7a7	Amino acid catabolism	Cationic AA transporter	Lysinuric Protein Intolerance	
Slc1a4	Amino acid catabolism	High-affinity glutamate/neutral AA	Intellectual Disability & Microcephaly	
Slc15a3	Amino acid catabolism	Proton oligopeptide		
Slc17a5	Amino acid catabolism	Vesicular glutamate	Sialic Acid Storage Diseases. Salla Disease	
SLC36a1	Amino acid catabolism	Proton-coupled AA		
Slc2a1	Carbohydrate metabolism	Facilitative GLUT	GLUT1 Deficiency Syndrome and Epilepsy	
Slc20a1	Carbohydrate metabolism	Phosphate		Combined pituitary hormone defic.
Slc41a2	Carbohydrate metabolism	MgtE-like magnesium		
Slc6a6	Cell membrane stability & volume regulation	Na & Cl-dependent Taurine		Retinal Degeneration
Slc12a2	Cell membrane stability & volume regulation	Electroneutral cation-coupled Cl	Child presenting w/ multiorgan dysfunction	Schizophrenia, Salivation Deficiency, Deafness, Male infertility
Slc14a1	Cell membrane stability & volume regulation	Urea	Chronic Mulifocal Osteomyelitis	Human Urothelial Cancer
Slc39a10	Cell membrane stability & volume regulation	Metal ion		
Slc25a42	FAO, OXPHOS	Mitochondrial carrier	Mitoch. Myopathy/Encephalomyelopathy	
Slc25a45	FAO, OXPHOS	Mitochondrial carrier		
Slc33a1	FAO, OXPHOS	Acetyl-CoA	Huppke-Brendel Syndrome, Autosomal Dominant Spastic Paraplegias	Autism Spectrum Disorder
Slc41a3	FAO, OXPHOS	MgtE-like magnesium		Hypomagnesemia, Hydronephrosis
Slc45a1	FAO, OXPHOS	Putative sugar	Intellectual Disability and Epilepsy	
Slc29a3	FAO, OXPHOS	Facilitative nucleoside	H syndrome, Pigmented Hypertrichotic Dermatosis w/ Diabetes, Dysosteosclerosis	
Slc31a2	FAO, OXPHOS	Copper		
Slc35a4	Glycosylation	Nucleoside-sugar		
Slc35e1	Glycosylation	Nucleoside-sugar		
Slc35d2	Glycosylation	Nucleoside-sugar		
Slc35f5	Glycosylation	Nucleoside-sugar		
Slc4a7	Intracellular pH	Bicarbonate		Breast Cancer, Blood Pressure
Slc26a6	Intracellular pH	Multifunctional anion		IBD, Crohn's, Ulcerative Colitis
Slc16a1	Intracellular pH	Monocarboxylate	Recurrent Ketoacidosis	Hyperinsulinism
Slc29a1	Nucleoside salvage	Facilitative nucleoside	Augustine Blood Group	Mineralization
Slc29a2	Nucleoside salvage	Facilitative nucleoside		
Slc5a6	Vitamin transport	Sodium glucose	Child presenting w/ multiorgan dysfunction	
Slc19a1	Vitamin transport	Folate/thiamine		Pregnancy Loss, Myelomeningocele
Slc46a1	Vitamin transport	Folate	Hereditary Folate Malabsorption	
Slc6a4	Serotonin transport	Na & Cl-dependent neurotransmitter		
Slc45a4	Unknown	Putative sugar		Multiple Blood Cell Variants

Extended Data Fig. 2 | Regulation of SLC expression during efferocytosis. a, SLC genes are differentially regulated during efferocytosis. Left, plot of the 165 SLC genes detected by RNA-seq of efferoytic LR73 cells, highlighting the 19 significantly upregulated (red) and 14 downregulated (blue) SLC genes that were altered during efferocytosis. The 132 SLC genes that were not altered are located on the midline (black). Right, the current genetic classifications of these 33 SLC genes that are altered during engulfment are shown. **b**, Efferocytosis-associated SLCs and their properties. Current genetic classification and/ or functional linkages of the 33 SLCs modulated during apoptotic cell engulfment. The significantly upregulated and downregulated SLCs and the substrates they are known to transport grouped by predicted general function are shown, as are the known monogenic diseases and single nucleotide polymorphism (SNP) or disease phenotype to which the specific SLCs have been linked.

a

**OXPHOS**   **A.A. Transport & Catabolism**   **Glycosylation**
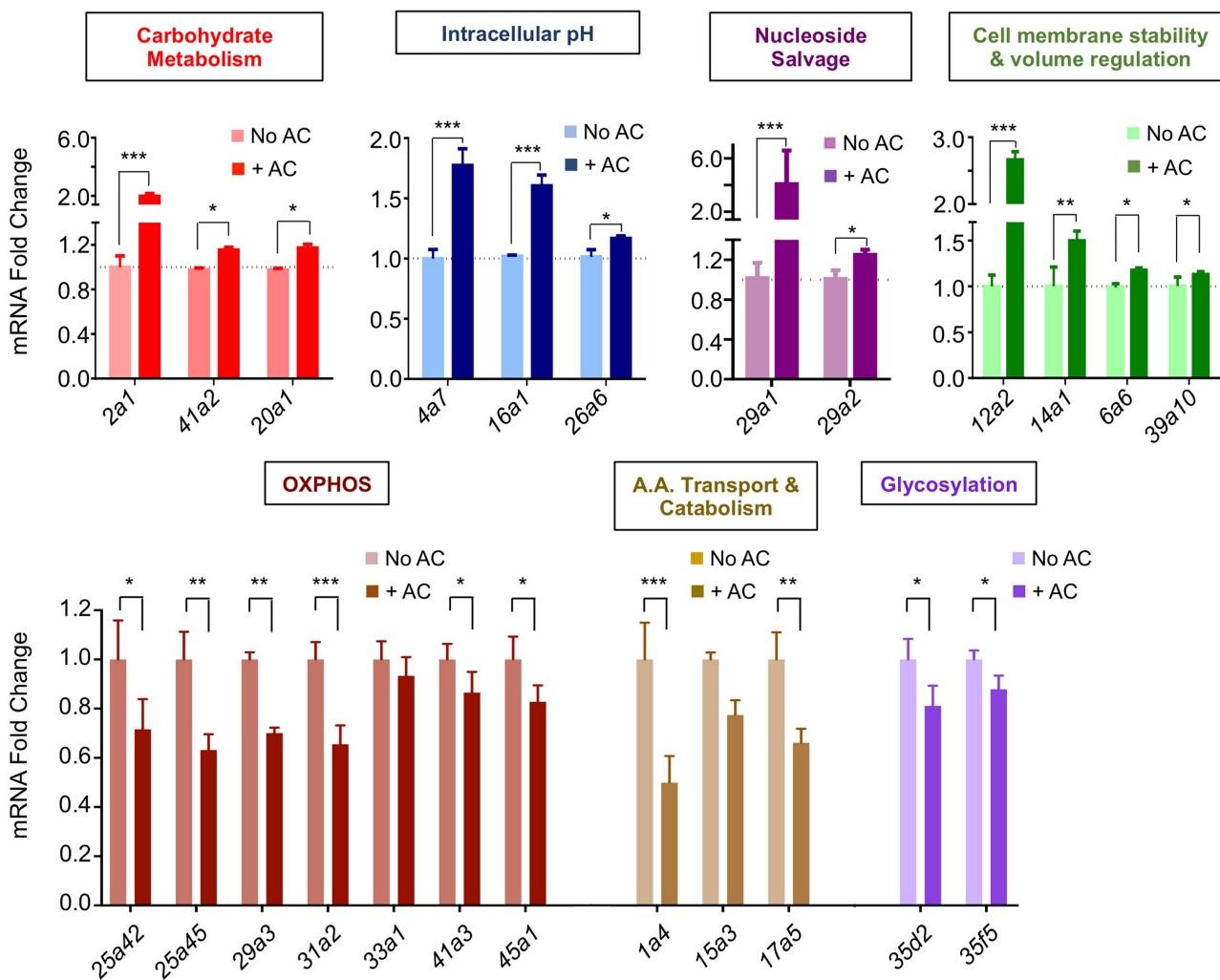
b

Primer Species	TaqMan Primer	Cycle number product detected in:			Primer Species	TaqMan Primer	Cycle number product detected in:		
		LR73 (Hamster)	Macrophage (Mouse)	Jurkat (Human)			LR73 (Hamster)	Macrophage (Mouse)	Jurkat (Human)
Hamster	2a1	26		Undetected	Mouse	2a1		25	Undetected
	41a2	27		Undetected		20a1		27	Undetected
	20a1	23		Undetected		14a1		30	Undetected
	4a7	32		Undetected		12a2		29	Undetected
	16a1	25		Undetected		29a1		27	Undetected
	26a6	27		Undetected		29a2		31	Undetected
	29a1	25		Undetected		4a7		27	Undetected
	29a2	26		Undetected		16a1		27	Undetected
	12a2	27		Undetected		29a3		7	Undetected
	14a1	30		Undetected		Gapdh		20	Undetected
	6a6	25		Undetected					
	39a10	26		Undetected					
	Gapdh	17		Undetected					

Extended Data Fig. 3 | qRT–PCR confirmation of the RNA-seq data. **a**, qRT–PCR of mRNA of specific SLCs during efferocytosis. Indicated SLC genes were tested for mRNA expression levels during engulfment assays performed similarly to those in Fig. 1a. Data are representative of at least two independent experiments with 3–4 replicates per condition. **b**, The table presents the cycle numbers for each species-specific qRT–PCR primer. None of these primers produced signals when tested against human Jurkat cell mRNA (target) alone.

a

b

c

d
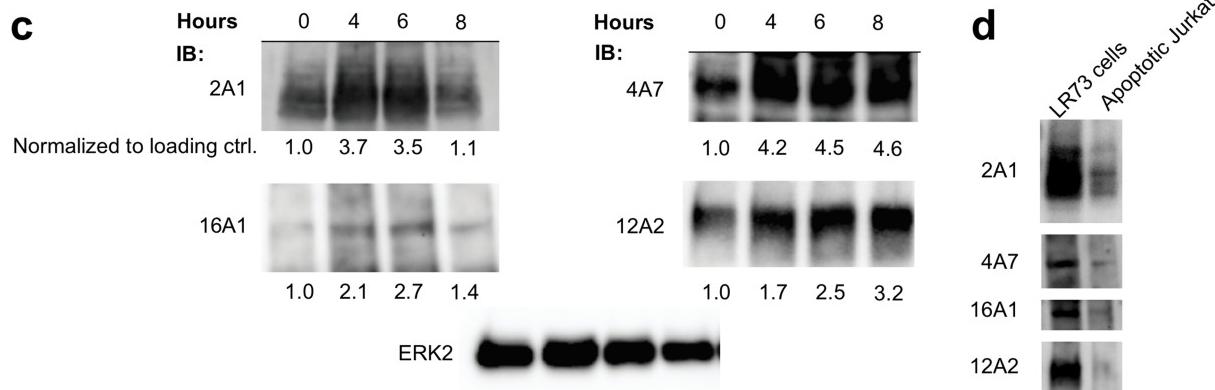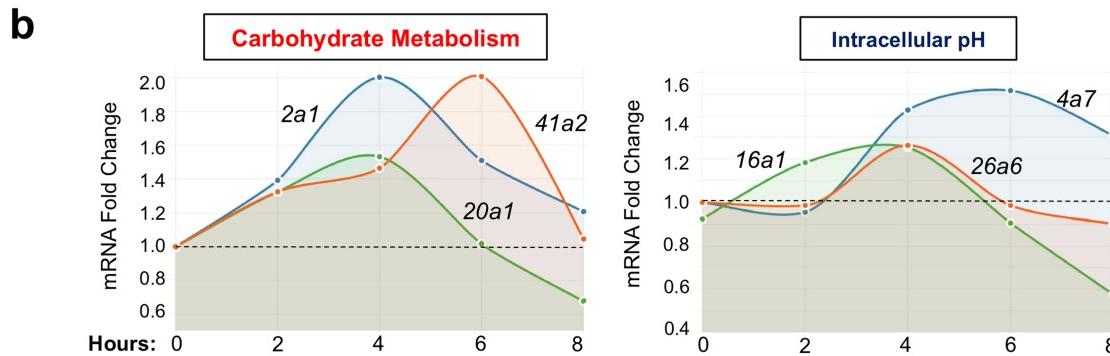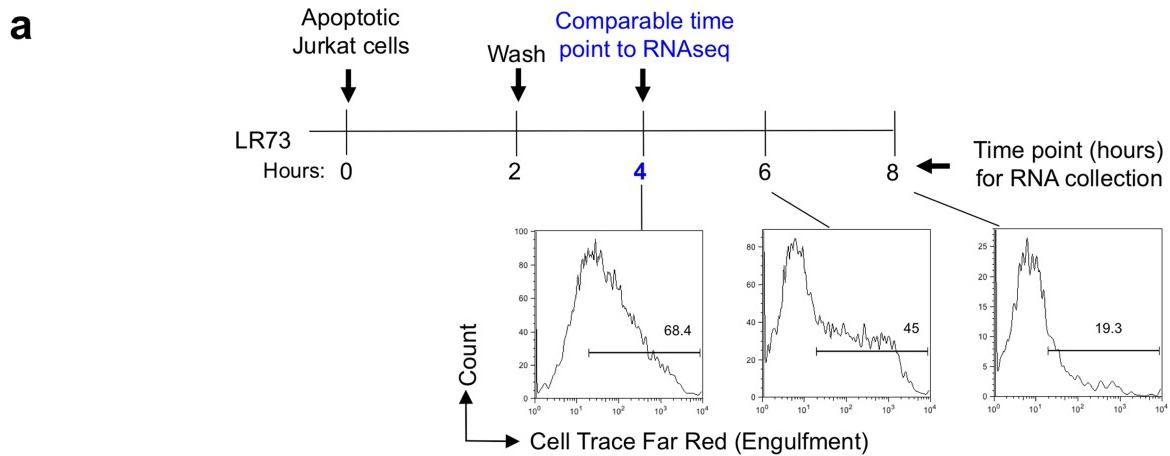
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Dynamic expression of SLCs during efferocytosis. a, Schematic of the experiment and time points when RNA from phagocytes was assessed for specific SLC gene expression. Apoptotic Jurkat cells were added to LR73 cells and co-cultured for 2 h. Unbound and floating apoptotic cells were then washed away, and the LR73 cells were cultured in fresh medium for the indicated times. The time scale bar reflects total time of experiment, such that the 4-h time point reflects 2 h with apoptotic cells plus 2 h subsequent incubation (to match the timeframe used in our RNA-seq experiment). Total RNA was subsequently isolated and qRT–PCR was performed for specific SLC genes. Flow cytometry plots indicate that fluorescent signals from the internalized corpses are significantly degraded by the 8 h time point. **b**, Expression of SLC genes is regulated over the time course of efferocytosis. Relative expression of mRNAs for specific SLC genes belonging to different functional classes over the time course of engulfment is shown. Data are representative of three biological replicates. **c**, Immunoblotting for some of the SLCs modified during efferocytosis. Indicated SLCs were probed at various time points after addition of apoptotic cells. Relative intensities of specific bands, normalized to ERK2, are shown below representative blots. **d**, Immunoblotting for the some of the SLCs in LR73 phagocytes and apoptotic Jurkat cells.

a Normal Antibody-mediated phagocytosis by SLC2A1-deficient BMDMs

b *Slc2a1* siRNA knockdown and rescue

c Confirmation of *Slc2a1* deletion in LR73 cells by CRISPR/Cas9 (qPCR)


**d** Reduction of SLC2A1 protein in BMDMs from *Slc2a1*^fl/fl^ mice

e *In vivo* antibody-mediated phagocytosis is not blocked by STF-31

f STF-31 does not further reduce engulfment in *Slc2a1* deficient cells

g Apoptosis induction in thymocytes after dex treatment is not altered by STF-31


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | The role of SLC2A1 in efferocytosis. a, *Slc2a1*^{fl/fl} BMDMs were treated with or without TAT-Cre to delete *Slc2a1*. The cells were then incubated with IgG-coated Jurkat cells and engulfment was assessed by CypHer5E signal within BMDMs. The uptake by control BMDMs (not treated with TAT-Cre, and denoted wild type (WT)) was set to 1. **b**, siRNA targeting of *Slc2a1* downregulates SLC2A1 protein expression. Representative western blots from siRNA knockdown of *Slc2a1* versus scrambled siRNA in LR73 cells are shown. LR73 cells expressing siRNA-resistant SLC2A1 are also shown. **c**, *Slc2a1* deletion efficiency in Cas9 LR73 cells. *Slc2a1* guide was introduced into Cas9–EGFP⁺ LR73 cell clones. The efficiency of *Slc2a1* deletion was quantified using qRT–PCR. **d**, Introduction of TAT-Cre into *Slc2a1*^{fl/fl} BMDMs efficiently knocks down SLC2A1 protein expression. *Slc2a1*^{fl/fl} bone marrow cells were treated with recombinant TAT-Cre during macrophage differentiation after isolation from the bone marrow. **e**, STF-31 did not affect antibody-mediated phagocytosis by peritoneal macrophages. C57BL/6 mice were intraperitoneally injected with 10 mg kg^{−1} of either

STF-31 in X-VIVO medium 1 h before injection of IgG-coated Jurkat cells. CypHer5E-labelled Jurkat cells were injected intraperitoneally along with the drug. Mice were euthanized 1 h later, peritoneal cells were collected, and apoptotic cell engulfment by CD11b⁺F4/80^{high} macrophages was analysed by fluorescence-activated cell sorting. **f**, *Slc2a1*-deficient LR73 cells or BMDMs were treated with STF-31, and the engulfment assay was conducted using CypHer5E-labelled apoptotic Jurkat cells. CypHer5E⁺ phagocytic cells after 2 h of incubation were identified by flow cytometry. n.s., not significant. Data are representative of at least two independent experiments with 3–4 replicates per condition. **g**, The SLC2A1 inhibitor STF-31 does not increase the number of thymocytes stained with 7-aminoactinomycin D (7AAD⁺) in vitro. Isolated thymocytes were incubated with dexamethasone (10 μM) with or without STF-31 (2 mM). Four hours later, the cell death of the thymocytes was indicated by annexin7⁺7AAD⁺. Data are representative of two independent experiments.

a Glucose promotes efferocytosis via SLC2A1

b *In vitro* efferocytosis with 2-DG
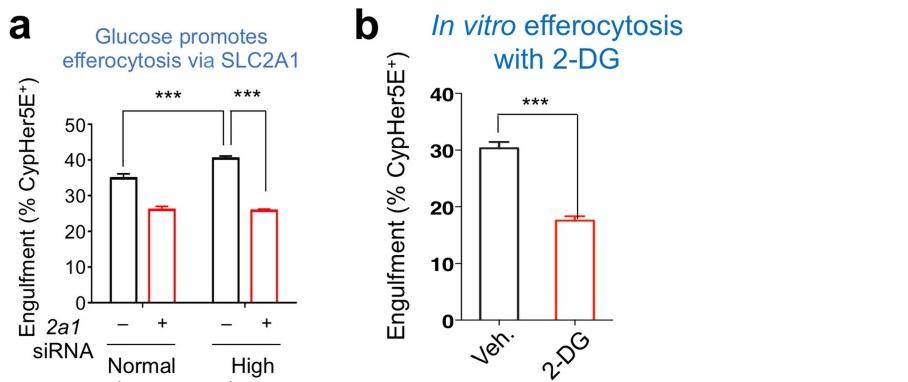
c Seahorse analysis of bone marrow-derived macrophages during efferocytosis

Glycolysis

OXPHOS

d
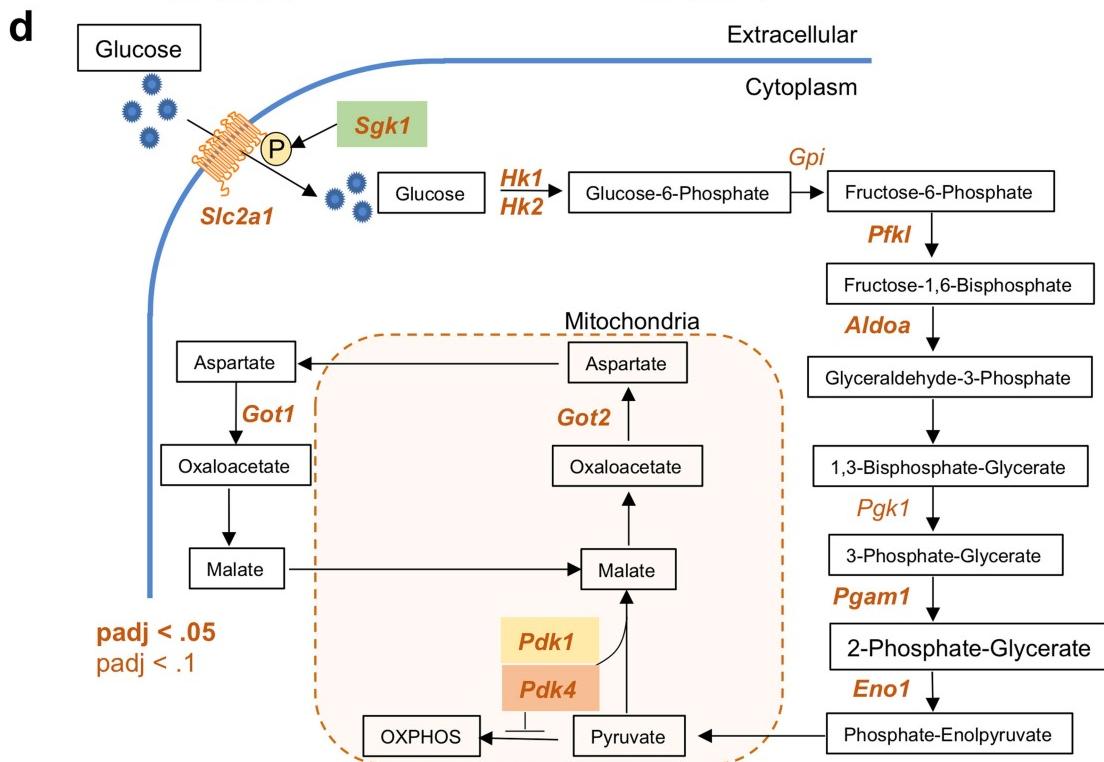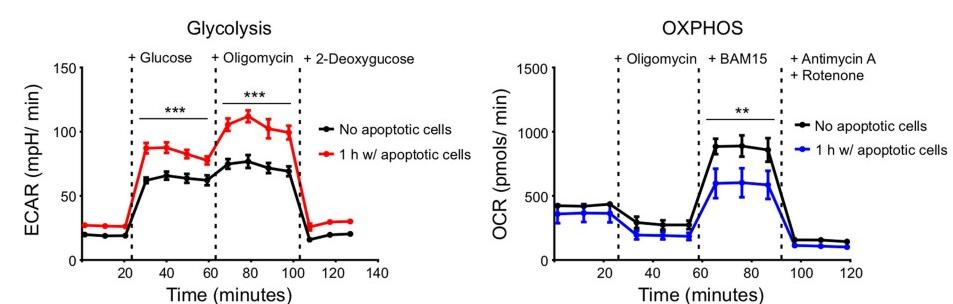
Genes indicated in red- Glycolytic pathway intermediates upregulated during efferocytosis

**Extended Data Fig. 6 | The role of glycolytic genes in efferocytosis.**
a, The effect of physiological (1 mg ml^{-1}) or high (5 mg ml^{-1}) glucose on apoptotic cell engulfment (2 h) in control and *Slc2a1*-siRNA-treated LR73 cells. Note that the enhanced engulfment due to higher glucose concentration is lost in siRNA-treated conditions. Data are representative of at least three independent experiments with 3–4 replicates per condition. **b**, Apoptotic cell engulfment by LR73 cells in the presence of the glucose analogue 2-DG (10 mM). Data are representative of two independent experiments with 2–3 replicates per condition. **c**, BMDMs undergo glycolytic flux during apoptotic cell clearance. Glycolytic flux and OXPHOS were measured during engulfment assays using Seahorse XF to assess ECAR (left) and OCR (right), respectively. Data are mean ± s.d. for ECAR and OCR over the course of standard glycolytic flux and cellular respiration tests. Data are representative of four replicates per condition. **d**, Genes within the glycolytic pathway that are significantly upregulated during apoptotic cell clearance. A schematic of the glycolytic pathway and subsequent steps is shown, with enzymes that are significantly upregulated (determined via RNA-seq) indicated in red.

Extended Data Fig. 7 | Testing SGK1 and glycolysis in efferocytosis.
a, Differential metabolic requirements of macrophages for efferocytosis versus antibody-mediated phagocytosis. BMDMs were co-cultured with apoptotic or antibody-coated Jurkat cells. Mitochondrial respiration was inhibited by addition of the mitochondrial complex I inhibitor rotenone (200 nM), the mitochondrial complex III inhibitor antimycin A1 (1 μM), or both (R + A). Aerobic glycolysis was inhibited by the addition of the pan-PDK inhibitor dichloroacetate (1 mM). Data are representative of three independent experiments. **b**, SGK1 inhibition blocks efferocytosis in vitro. LR73 cells were treated with SGK1 inhibitor and uptake of CypHer5E-labelled apoptotic Jurkat cells was assessed. **c**, BMDMs from GLUT1–Myc knock-in mice were co-cultured with apoptotic thymocytes with or without the SGK1 inhibitor GSK650394 (5 μM) for 2 h, unbound apoptotic cells were washed away and the cell-surface expression of SLC2A1 was measured by flow cytometry after staining for surface Myc tag. Data are representative of at least two independent experiments. **d**, Continued uptake of apoptotic thymocytes was determined by the MFI (indicative of corpse-derived signal per phagocyte) of LR73 phagocytes over a time course of engulfment. SLC2A1 or SGK1 inhibitors were added at the beginning of engulfment (left of each pair of graphs) or 1 h post-apoptotic cell addition (right of each pair of graphs). Data are representative of at least three independent experiments with 3–4 replicates per condition.

a ATP-induced *Sgk1* expression

b PtdSer masking via GST-TSR inhibits upregulation of Slc2a1 and Slc16a1, but not Sgk1

c siRNA knockdown of *Slc16a1* affects engulfment by LR73 cells

d Reduction of engulfment by SR13800 in CD11b⁺ F4/80⁺ peritoneal macrophages *in vivo*

e Pro-inflammatory markers qPCR

☐ Sup from LR73 alone   ☐ Sup from LR73 + ACs

f CD206 expression after supernatant treatment of BMDMs
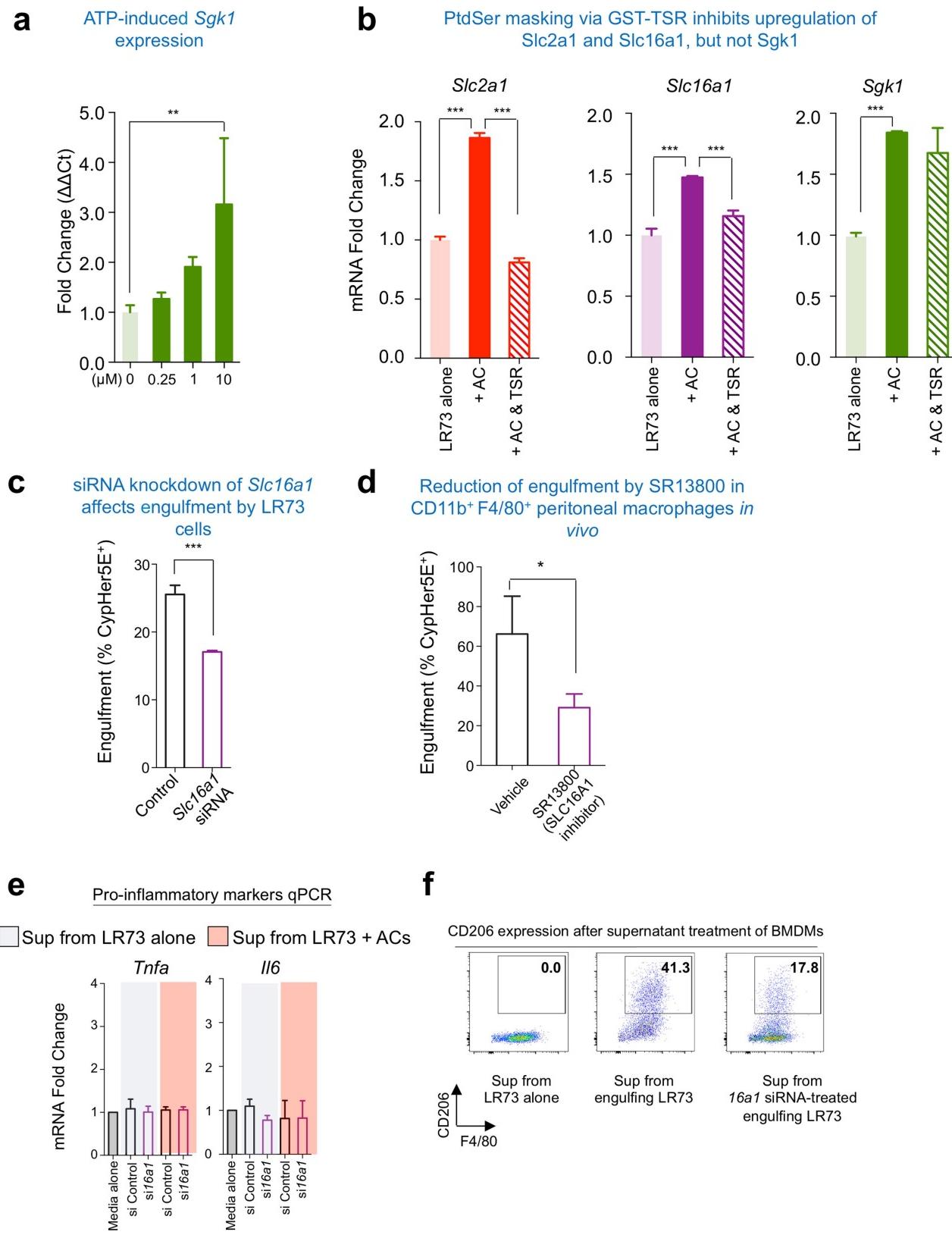
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Testing SLC16A1 in efferocytosis. a, qRT–PCR determination of *Sgk1* expression in phagocytes treated with ATP. LR73 cells were treated with indicated amounts of ATP for 4 h. Expression of *Sgk1* was determined by qRT–PCR using hamster-specific primers. Data are representative of at least two independent experiments with 3–4 replicates per condition. **b**, qRT–PCR determination of *Slc2a1*, *Slc16a1* and *Sgk1* expression in phagocytes after addition of the PtdSer-masking peptide (GST–TSR) during efferocytosis. Apoptotic cells (AC) were added with or without TSR peptide (10 ng μl^{-1}) for 4 h. Expression of indicated genes was determined by qRT–PCR using hamster-specific primers. Data are representative of at least two independent experiments with 3–4 replicates per condition. **c**, SLC16A1 inhibition blocks efferocytosis in vitro. LR73 cells were treated with *Slc16a1* siRNA and uptake of CypHer5E-labelled apoptotic Jurkat cells was assessed.

d, SLC16A1 inhibitor SR13800 dampens efferocytosis by peritoneal macrophages. C57BL/6 mice were injected intraperitoneally with SR13800 (10 mg kg^{-1}) in X-VIVO medium 1 h before injection of apoptotic cells. CypHer5E-labelled apoptotic Jurkat cells were injected intraperitoneally. After 1 h, apoptotic cell engulfment by CD11b^{+}F4/80^{high} peritoneal macrophages was analysed by flow cytometry. Data are representative of two independent experiments with at least six mice in each group per experiment. **e, f**, Supernatants were prepared from LR73 cells, treated with control or *Slc16a1* siRNA, that were engulfing apoptotic cells. The supernatants were added to BMDMs and incubated for 12 h. **e**, Expression of inflammatory markers was determined by qRT–PCR. **f**, After 24 h of incubation, expression of CD206 and F4/80 was determined by flow cytometry. Data are representative of two independent experiments with 2–3 replicates per condition.

Mannose impairs tumour growth and enhances chemotherapy

Pablo Sierra Gonzalez[1], James O'Prey[1], Simone Cardaci[1,2,6], Valentin J. A. Barthet[1,6], Jun-ichi Sakamaki[1], Florian Beaumatin[1], Antonia Roseweir[3], David M. Gay[1], Gillian Mackay[1], Gaurav Malviya[1], Elżbieta Kania[1], Shona Ritchie[1], Alice D. Baudot[1], Barbara Zunino[1], Agata Mrowinska[1], Colin Nixon[1], Darren Ennis[3,5], Aoisha Hoyle[4], David Millan[4], Iain A. McNeish[3,5], Owen J. Sansom[1,3], Joanne Edwards[3] & Kevin M. Ryan[1,3]*

It is now well established that tumours undergo changes in cellular metabolism[1]. As this can reveal tumour cell vulnerabilities and because many tumours exhibit enhanced glucose uptake[2], we have been interested in how tumour cells respond to different forms of sugar. Here we report that the monosaccharide mannose causes growth retardation in several tumour types in vitro, and enhances cell death in response to major forms of chemotherapy. We then show that these effects also occur in vivo in mice following the oral administration of mannose, without significantly affecting the weight and health of the animals. Mechanistically, mannose is taken up by the same transporter(s) as glucose[3] but accumulates as mannose-6-phosphate in cells, and this impairs the further metabolism of glucose in glycolysis, the tricarboxylic acid cycle, the pentose phosphate pathway and glycan synthesis. As a result, the administration of mannose in combination with conventional chemotherapy affects levels of anti-apoptotic proteins of the Bcl-2 family, leading to sensitization to cell death. Finally we show that susceptibility to mannose is dependent on the levels of phosphomannose isomerase (PMI). Cells with low levels of PMI are sensitive to mannose, whereas cells with high levels are resistant, but can be made sensitive by RNA-interference-mediated depletion of the enzyme. In addition, we use tissue microarrays to show that PMI levels also vary greatly between different patients and different tumour types, indicating that PMI levels could be used as a biomarker to direct the successful administration of mannose. We consider that the administration of mannose could be a simple, safe and selective therapy in the treatment of cancer, and could be applicable to multiple tumour types.

As tumours often have a high avidity for glucose[4], we examined the effect of other sugars on the growth of tumour cells. This revealed that mannose, in contrast to other sugars, significantly reduced the growth of U2OS cells (Fig. 1a). Similar effects occurred in Saos-2 cells, with fucose also causing a comparatively small decrease in cell growth (Extended Data Fig. 1a). Using a panel of cell lines, we observed that this effect of mannose occurs in cells from various tissues, with the effect being greater in some cells than others (Fig. 1b, c, Extended Data Fig. 1a–d).

Mannose is imported into cells by means of the same transporters as glucose[3], so we considered that mannose might interfere with the uptake of glucose. In support of this hypothesis, we observed that mannose enhances levels of phosphorylated AMPK—a read-out of energy balance in cells[5,6] (Extended Data Fig. 1e). However, liquid chromatography coupled with mass spectrometry (LC–MS) analyses showed that levels of the phosphorylated form of the glucose analogue 2-deoxyglucose (2-DG-P)—used as a proxy for glucose uptake—did not correlate with mannose sensitivity in mannose-sensitive compared with mannose-insensitive cell lines (Fig. 1a, b, Extended Data Fig. 1f, g). In fact, mannose increased the intracellular pool of hexose-6-phosphate, which is

produced in the first step of the metabolism of glucose and mannose (Extended Data Fig. 1h). This effect was not observed with other sugars (Extended Data Fig. 1i).

To determine whether the mannose-induced increase in hexose-6-phosphate levels is due to mannose or glucose, we performed LC–MS with an extended chromatography time to enable selective detection of the unphosphorylated form of these sugars. This revealed that mannose increased the intracellular pool of mannose, as expected, but it also increased the intracellular pool of glucose (Extended Data Fig. 1j, k). To corroborate this finding, we treated cells with glucose labelled with two atoms of ^{13}C (1,2-$^{13}C_2$-glucose) and mannose labelled with six atoms of ^{13}C ($^{13}C_6$-mannose) to enable detection due to differences in molecular mass; we again found that treatment with mannose increased intracellular levels of glucose (Extended Data Fig. 1l).

As mannose did not reduce intracellular levels of glucose, but significantly affected cell growth, we considered that it might interfere with glucose metabolism. There is considerable crosstalk between the metabolism of these sugars, and mannose-6-phosphate can inhibit three enzymes that mediate glucose metabolism: hexokinases, phosphoglucose isomerase (PGI) and glucose-6-phosphate dehydrogenase[7] (Fig. 1d). To address this possibility, we measured the levels of hexose-6-phosphate and lactate when cells were incubated in glucose-free medium supplemented with either 1,2-$^{13}C_2$-glucose or $^{13}C_6$-mannose. This revealed that mannose accounts for more of the mannose-induced hexose-6-phosphate accumulation than does glucose, but also that glucose-6-phosphate is produced in the presence of mannose, albeit at lower levels when compared with cells treated with glucose alone (Fig. 1e). We also observed that mannose produces only small amounts of lactate—indicating that it is poorly metabolised—and more notably, mannose markedly reduces the production of lactate from glucose (Fig. 1f). Further analysis revealed that mannose treatment not only affects pools of glycolytic intermediates, but also affects those involved in the tricarboxylic acid cycle, the pentose phosphate pathway and glycan synthesis (Fig. 1g–i). As with the effects on cell growth, these metabolic effects were not observed to the same extent with other sugars (Extended Data Fig. 1m–p). Moreover, although mannose uptake was not lower in mannose-insensitive cells when compared to mannose-sensitive cells (Extended Data Fig. 2a), the effects of mannose on metabolism were only observed in cells sensitive to the sugar (Extended Data Fig. 2b–h). We also found that mannose-induced AMPK phosphorylation does not occur concomitantly with changes in the levels of AMP or ATP, but is associated with a decrease in fructose-1,6-bisphosphate levels, as has been recently described[6] (Extended Data Fig. 3a–f).

Because mannose affects the growth of tumour cells, we questioned whether it can also affect the cellular response to chemotherapeutic drugs. Although mannose alone did not affect cell viability, it significantly enhanced cell death when administered with cisplatin or doxorubicin—an effect not seen with other hexoses (Fig. 2a, b,
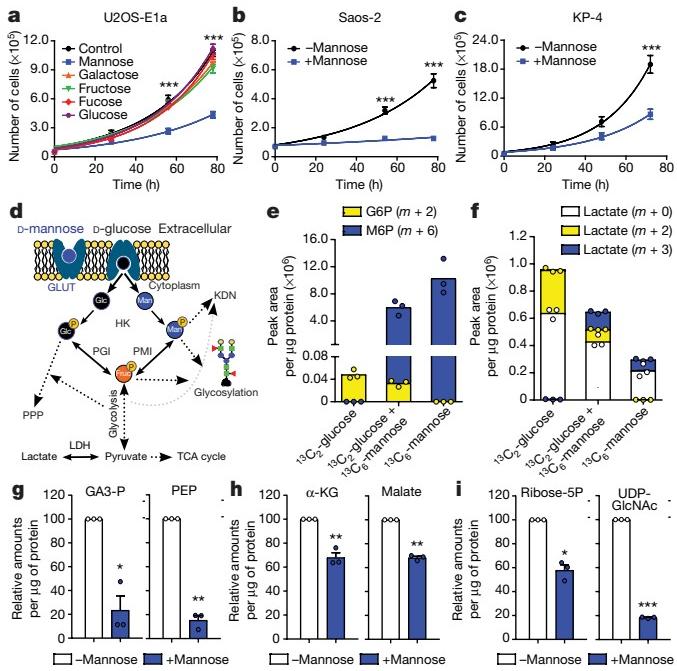
Fig. 1 | Mannose impairs the growth of cancer cells and interferes with glucose metabolism by accumulating intracellularly as mannose-6-phosphate. a, Growth curves of U2OS-E1a cells supplemented without (control) or with 25 mM of the hexoses stated. **b, c**, Growth curves of Saos-2 cells in DMEM alone (−mannose) or DMEM with an additional 25 mM mannose (+mannose) (**b**) and KP-4 cells in IMDM alone or with 25 mM mannose (**c**). **d**, Scheme of mannose metabolism: mannose enters the cells using the same transporters as glucose (GLUT) and is phosphorylated into mannose-6-phosphate by hexokinases (HK). Mannose can then be used for glycosylation purposes or isomerized into fructose-6-phosphate by PMI; both PMI and PGI can also produce mannose-6-phosphate from glucose-6-phosphate. Mannose-6-phosphate also participates in the biosynthesis of deaminoneuraminic acid (KDN). Fruc, fructose; glc, glucose; man, mannose; TCA cycle, tricarboxylic acid cycle. **e, f**, Extraction of intracellular metabolites and measurement of the peak area per microgram of protein of glucose-6-phosphate (G6P) ($m + 2$) and mannose-6-phosphate (M6P) ($m + 6$) (**e**) and lactate ($m + 0$, $m + 2$, $m + 3$) (**f**), after 6-h incubation of U2OS-E1a cells in 10% dialysed FBS in glucose-free DMEM complete medium in the presence of 5 mM 1,2-$^{13}C_2$-D-glucose alone, 5 mM $^{13}C_6$-D-mannose alone, or both sugars in combination. **g–i**, Relative amounts per microgram of protein of the intracellular metabolites glyceraldehyde-3-phosphate (GA3-P) (left) and phosphoenolpyruvate (PEP) (right) (**g**); α-ketoglutarate (α-KG) (left) and malate (right) (**h**); ribose-5-phosphate (ribose-5P) (left) and UDP-N-acetyl-glucosamine (UDP-GlcNAc) (right) (**i**) after a 6-h incubation of U2OS-E1a cells in 10% dialysed FBS in DMEM complete medium with 5 mM glucose, with or without 5 mM mannose as indicated. $n = 3$ independent experiments (**a–c, g–i**); data are representative of three independent experiments (**e, f**). Data are mean ± s.e.m. and were analysed by two-way ANOVA with Bonferroni correction (**a–c**) or paired two-tailed Student's t-test (**g–i**). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Extended Data Fig. 4a–c). Mechanistically, the combination of mannose and the chemotherapeutic drug increased the levels of cleaved poly(ADP-ribose) polymerase (PARP, a substrate[8] of caspase-3) and this effect, together with cell death, was blocked by the pan-caspase inhibitor zVAD-FMK (Fig. 2c, d, Extended Data Fig. 4d).

Because our data indicated that cell death after treatment with mannose and the chemotherapeutic drug is likely to proceed by apoptosis, we used CRISPR–Cas9 to determine which apoptotic pathways might be involved. Disruption of caspase 8 and FADD—two components of the extrinsic apoptotic pathway[9]—had no effect on cell death (Extended Data Fig. 4e–g). By contrast, disruption of Bax and Bak—essential factors for mitochondrial outer membrane permeabilization and the intrinsic pathway[10]—markedly reduced cell death after treatment with

mannose in combination with either cisplatin or doxorubicin (Fig. 2e, Extended Data Fig. 4h, i).

As mitochondrial outer membrane permeabilization is controlled by members of the Bcl-2 family[11], we examined the levels of these proteins in the presence of mannose, cisplatin or both in combination. In agreement with previous studies that detail a role for the Noxa/Mcl-1 axis in cell death induced by glycolysis inhibition and glucose deprivation[12–14], we found that levels of Noxa increased after treatment with cisplatin and cisplatin plus mannose, and levels of Mcl-1 and Bcl-X_L decreased after the combination treatment (Fig. 2f, Extended Data Fig. 4j). Mechanistically, we consider that the change in Mcl-1 and Bcl-X_L levels is due to decreased translation because the levels of the proteins themselves were still decreased in the presence of a proteasome inhibitor, and there was no change in the levels of their mRNAs, after the combination treatment (Extended Data Fig. 4k–m). The involvement of these proteins was confirmed by CRISPR-mediated disruption of *PMAIP1* (also known as *NOXA*) or by overexpression of Mcl-1 or Bcl-X_L, which all suppressed cell death induced by mannose and chemotherapy (Fig. 2g, h, Extended Data Figs. 4n–q, 5a). Finally, we also observed that mannose enhances cell death induced by the Bcl-X_L antagonist WEHI539, but not the Bcl-2 antagonist ABT-199 (Extended Data Fig. 5b).

We were keen to find out whether mannose also has effects in vivo. Tumour-bearing mice were given a single oral gavage of mannose, which resulted in a serum concentration of approximately 3 mM (Extended Data Fig. 6a). Mice were then injected with 2-deoxy-2-[^{18}F] fluoro-D-glucose ([^{18}F]FDG) to monitor [^{18}F]FDG uptake and its subsequent conversion by hexokinases into [^{18}F]FDG-6-phosphate[15]. As mannose impedes hexokinases[6], we found that the [^{18}F]FDG signal (provided by [^{18}F]FDG and/or [^{18}F]FDG-6-phosphate) was significantly reduced in mannose-treated mice when compared with control mice bearing tumours of an equivalent size (Fig. 3a, Extended Data Fig. 6b, c). A significant effect of mannose treatment on the [^{18}F]FDG signal was also seen in certain normal tissues (Extended Data Fig. 6d).

To test whether mannose could affect tumour growth, mice were injected with tumour cells subcutaneously and were given mannose both freely in drinking water and three times a week by oral gavage. This did not affect the weight of the mice, nor did it visibly affect their health (Extended Data Fig. 6e). However, it significantly inhibited tumour growth (Fig. 3b), involving decreased numbers of BrdU-positive cells, which indicates that mannose inhibits cell proliferation both in vitro and in vivo (Extended Data Fig. 6f, g).

We were also interested to know whether mannose can enhance chemotherapy in vivo. Tumour-bearing nude mice were treated with mannose and doxorubicin either alone or in combination. Although none of the treatments affected the weight or visibly affected the health of the mice (Extended Data Fig. 6h), we found that either doxorubicin or mannose caused a reduction in tumour volume (Fig. 3c). Moreover, an even greater effect was observed when doxorubicin was administered in combination with mannose (Fig. 3c). Notably, when we examined the overall survival of the treated cohorts, those treated with doxorubicin plus mannose had a significantly increased life expectancy when compared to untreated mice or those treated with either doxorubicin or mannose alone (Fig. 3d).

As our data indicated that mannose could potentially be used clinically, we wanted to understand why different cells vary in their sensitivity to the sugar. Because mannose affects several metabolic pathways, we reasoned that sensitivity may be connected to an apical enzyme involved in sugar metabolism. Our profiling revealed that mannose sensitivity was roughly inversely correlated with the levels and activity of PMI, the enzyme that catalyses the interconversion of mannose-6-phosphate and fructose-6-phosphate[16] (Figs. 1a–c and 4a, Extended Data Figs. 1b–d, 7a). Consequently, we knocked down *MPI*, the gene that encodes PMI, in three mannose-insensitive cell lines: SKOV3, RKO and IGROV1. In each case, *MPI* knockdown caused growth retardation upon mannose treatment (Fig. 4b, Extended Data Fig. 7b–g) and markedly sensitized cells to cell death when mannose
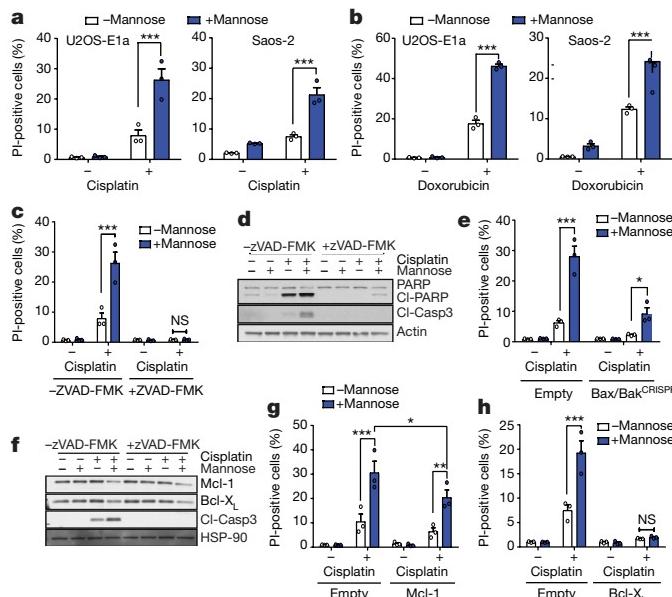
Fig. 2 | The combination of chemotherapeutic drugs with mannose enhances cell death by potentiating the intrinsic pathway of apoptosis through the downregulation of Mcl-1 and Bcl-X_L protein levels. a–h, All experiments were performed by pre-incubating cells in the presence of complete DMEM or complete DMEM supplemented with 25 mM sugars for 24 h before the addition of other treatments. a, b, The percentage of U2OS-E1a (left) and Saos-2 (right) propidium-iodide (PI)-positive cells after 24 h treatment with 10 μM cisplatin (a) and 1 μg ml^{-1} of doxorubicin (b) in the presence or absence of 25 mM mannose as indicated. c, The percentage of U2OS-E1a propidium-iodide-positive cells treated for 24 h with or without 10 μM cisplatin, with or without 25 mM mannose and with or without 50 μM of the caspase inhibitor zVAD-FMK. d, Western blots showing the levels of cleaved PARP (Cl-PARP) and cleaved caspase-3 (Cl-Casp3) in U2OS-E1a cells after 24 h treatment with or without 10 μM cisplatin, with or without 25 mM mannose and with or without 50 μM zVAD-FMK. e, The percentage of empty (left) and Bax/Bak^{CRISPR} (right) U2OS-E1a propidium-iodide-positive cells treated with or without 10 μM cisplatin and with or without 25 mM mannose for 24 h. f, Western blots showing the levels of Mcl-1, Bcl-X_L and cleaved caspase-3 in U2OS-E1a cells after 24 h with or without 10 μM cisplatin, with or without 25 mM mannose and with or without 50 μM zVAD-FMK. g, h, Percentage of propidium-iodide-positive U2OS-E1a (empty, Mcl-1 and Bcl-X_L overexpressing) cells after 24 h treatment with or without 10 μM cisplatin and with or without 25 mM mannose. $n = 3$ independent experiments (a–c, e, g, h); data are representative of three independent experiments (d, f). Data are mean ± s.e.m. and were analysed by two-way ANOVA with Bonferroni correction (a–c, e, g, h). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

was administered together with cisplatin (Fig. 4c). Conversely, the over-expression of *MPI* in a mannose-sensitive cell line rendered the cell line refractory to the effects of mannose on both cell growth and cell death (Fig. 4d, Extended Data Fig. 7h–k). Finally, we found that mannose treatment following *MPI* knockdown had highly significant effects on metabolic pathways downstream of the sugar (Fig. 4e, Extended Data Fig. 7l–o).

We next questioned whether PMI could be modulated to affect the response of tumours to mannose in vivo. As the immune microenvironment is important in many therapeutic situations and because mannose can affect immune cell function[17,18], we decided to use immune-competent mice and two syngeneic cell lines (B16-F1 and LLC) that are ordinarily mannose-insensitive but become mannose-sensitive upon *Mpi* knockdown (Fig. 4f, g, Extended Data Fig. 8a–d). In both cases, allografts formed with *Mpi*-knockdown cells were highly sensitive to the oral administration of mannose (Fig. 4h–k), without visibly affecting the health or weight of the mice (Extended Data Fig. 8e–h).
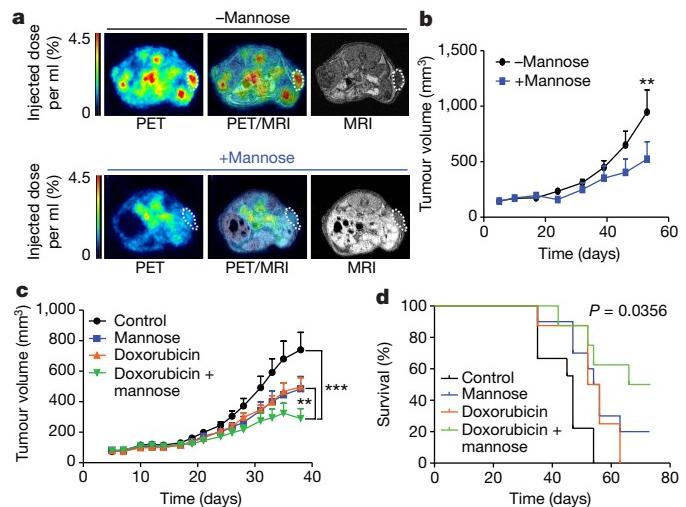

Fig. 3 | Mannose impairs tumour growth and induces tumour regression in combination with chemotherapy. a, CD1-nude mice were transplanted with KP-4 cells subcutaneously and tumours were grown for 14 days before positron emission tomography (PET) and magnetic resonance imaging (MRI) scans were carried out. a, PET and MRI scans of mice treated with 200 μl of water (top) or 200 μl of 20% (w/v) mannose in water (bottom) by oral gavage 20 min before injection of [^{18}F]FDG into the tail vein. White dotted circles highlight tumour areas in axial view of the mice. b, CD1-nude mice were injected with KP-4 cells subcutaneously and received either normal drinking water or 20% mannose in the drinking water, plus the same treatment (either normal water or 20% mannose) by oral gavage three days a week from the third day after tumour transplantation. Tumour volume (mm^3) was measured as indicated. c, d, CD1-nude mice were injected with KP-4 cells subcutaneously and tumours were grown for 10 days before the start of mannose treatment. Mice received either normal drinking water (control and doxorubicin) or 20% mannose in the drinking water (mannose and doxorubicin + mannose), plus the same treatment (either water or 20% mannose in water) via oral gavage three days a week. Doxorubicin treatment started on day 32 and mice received 5 mg kg^{-1} by intraperitoneal injection once a week. c, Tumour volume (mm^3) of all treatment groups. d, Graph representing the survival of the mice in all treatment groups until the end of the experiment at day 73. The number of mice for each experiment is as follows: $n = 5$ (−mannose), $n = 4$ (+mannose) (a); $n = 10$ (−mannose), $n = 8$ (+mannose) (b); $n = 10$ per group (c, d). Data are mean ± s.e.m. and were analysed by two-way ANOVA with Bonferroni correction (b, c) or log-rank test (two-sided Mantel–Cox test) (d). $**P < 0.01$, $***P < 0.001$.

We were keen to know whether PMI levels varied in human tumours, such that the analysis of PMI could potentially be used as a biomarker for mannose sensitivity. We therefore stained tissue microarrays containing sections from ovarian, renal, breast, prostate and colorectal cancers. This revealed that PMI levels not only varied among tumours from the same tissue, but also between tissues (Fig. 4l, Extended Data Fig. 9a). PMI levels did not, however, have prognostic significance in breast and colon cancer, presumably because normal serum levels of mannose are low compared to glucose[19] (Extended Data Fig. 9b–d).

Most notable from our analyses was the fact that colorectal tumours generally have very low PMI levels (Fig. 4l), which indicates that they may be broadly sensitive to mannose. To explore this, we used an inflammation-driven model of colorectal cancer and a genetically engineered mouse model driven by two genes that are frequently altered in this disease (*Kras* and *Apc*)[20]. In both models, mice maintained on drinking water containing 20% mannose had significantly fewer tumours at the clinical end point (Fig. 4m, n), and notably, mannose had no negative effect on the health or weight of the mice over the time examined (Extended Data Fig. 9e, f).

In contrast to its effects on glucose metabolism, mannose does not decrease the uptake of amino acids or fatty acids, and although mannose reduces glucose-dependent serine and glycine synthesis, this contributes only marginally to total cellular serine and glycine pools
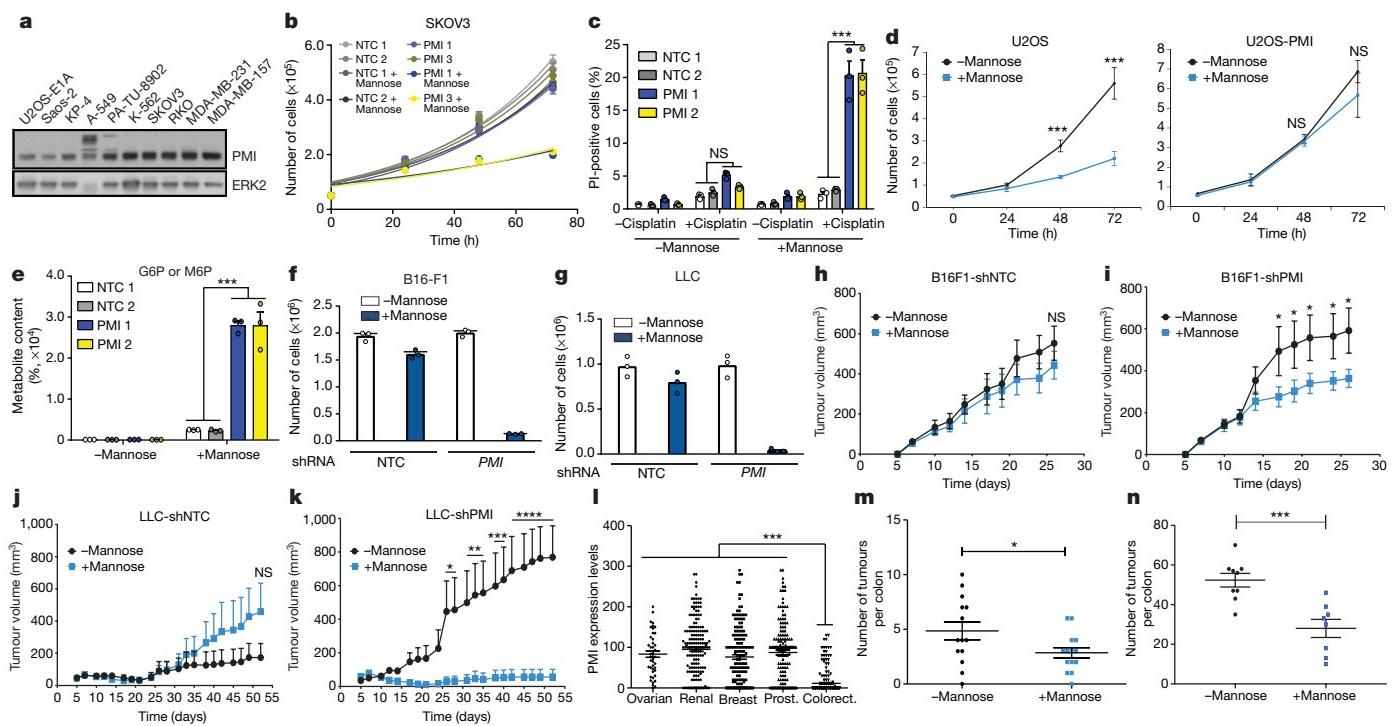
Fig. 4 | PMI levels dictate mannose sensitivity. a, Western blot showing the levels of PMI and ERK2 in a panel of ten cancer cell lines. **b**, Growth curves of SKOV3 cells in complete DMEM medium with or without supplementation of 25 mM mannose after transient transfection with two non-targeting (NTC) and two PMI-targeting short interfering RNAs (siRNAs) individually for 48 h. **c**, Cell death represented as the percentage of propidium-iodide-positive SKOV3 siRNA-transfected cells that had been pre-incubated with or without 25 mM mannose in regular DMEM for 24 h, before either 10 μM cisplatin was added or no cisplatin was added with incubation for a further 24 h. **d**, Overexpression of PMI renders U2OS-E1a insensitive to mannose. Growth curves of U2OS-E1a overexpressing PMI (U2OS-PMI) and U2OS-E1a cells expressing vector control (U2OS) after culture in either in DMEM or DMEM containing 25 mM mannose. **e**, Percentage of metabolite content of hexoses-6-phosphate in SKOV3 cells transfected with siRNA for 48 h before 6 h incubation in complete DMEM medium with or without supplementation of 25 mM mannose. Cells treated with NTC 1 without mannose were normalized to 100%. **f, g**, Knockdown of *Mpi* renders B16-F1 and LLC cells sensitive to mannose. The indicated cells were incubated with or without 25 mM mannose for 24 h before cell counting. **h–k**, B16-F1 or LLC cells expressing *Mpi*-targeting or control shRNAs were injected into the flanks of C57BL/6 mice. Mice were maintained either with or without 20% mannose in

drinking water and tumour growth was monitored over time ($n = 10$ mice per group). **l**, *MPI* expression levels in tissue microarrays from ovarian ($n = 45$), renal ($n = 180$), breast ($n = 159$), prostate (prost.) ($n = 155$) and colorectal (colorect.) ($n = 216$) tumours. **m**, Mice ($n = 14$ per group) were subjected to azoxymethane plus dextran sodium sulfate treatment for 68 days. Mice were treated with normal drinking water or with 20% mannose in drinking water until the clinical end point. Tumours were counted in the colon of each mouse. **n**, *Villin*^creER^*Apc*^fl/+^*Kras*^G12D/+^ mice were aged until the clinical end point. Mice were treated with normal drinking water or with 20% mannose in drinking water from four days post-induction until the clinical end point ($n = 9$ mice (−mannose), $n = 8$ mice (+mannose)). Tumours were counted in the colon of each mouse. $n = 3$ independent experiments (**b, c, e, f**); $n = 5$ independent experiments (**d**). Data represent one independent experiment performed in technical triplicate (**g**) or are representative of two independent experiments (**a**). Data are mean ± s.e.m. and are analysed by a one-sided Mann–Whitney *U*-test. ***$P < 0.001$. Data were analysed by two-way ANOVA with Bonferroni correction (**c, e**), multiple two-sided unpaired *t*-test with Holm–Sidak correction (**d**), two-way ANOVA with Tukey correction (**h–k**), one-way ANOVA with Bonferroni correction (**l**), unpaired two-tailed Student's *t*-test (**m**) and one-sided Mann–Whitney *U*-test (**n**). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

(Extended Data Fig. 10a–f). Mannose does not invoke an endoplasmic-reticulum stress response, but affects transcription, translation and autophagy, although these effects were reversed by overexpression of *MPI*, which indicates that they are downstream of glucose metabolism (Extended Data Fig. 10g–l). Moreover, the ablation of autophagy did not affect mannose sensitivity, showing that autophagy inhibition is not the mechanism underlying the effects of the sugar (Extended Data Fig. 10m, n). In summary, we conclude that mannose represents a well-tolerated means to interfere with glucose metabolism that could potentially be used clinically, either alone or in combination with other forms of cancer therapy.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0729-3.

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
2. Pavlova, N. N. & Thompson, C. B. The emerging hallmarks of cancer metabolism. *Cell Metab.* **23**, 27–47 (2016).
3. Thorens, B. & Mueckler, M. Glucose transporters in the 21st century. *Am. J. Physiol. Endocrinol. Metab.* **298**, E141–E145 (2010).
4. Cairns, R. A., Harris, I. S. & Mak, T. W. Regulation of cancer cell metabolism. *Nat. Rev. Cancer* **11**, 85–95 (2011).
5. Chaube, B. & Bhat, M. K. AMPK, a key regulator of metabolic/energy homeostasis and mitochondrial biogenesis in cancer cells. *Cell Death Dis.* **7**, e2044 (2016).
6. Zhang, C. S. et al. Fructose-1,6-bisphosphate and aldolase mediate glucose sensing by AMPK. *Nature* **548**, 112–116 (2017).
7. DeRossi, C. et al. Ablation of mouse phosphomannose isomerase (*Mpi*) causes mannose 6-phosphate accumulation, toxicity, and embryonic lethality. *J. Biol. Chem.* **281**, 5916–5927 (2006).
8. Fischer, U., Jänicke, R. U. & Schulze-Osthoff, K. Many cuts to ruin: a comprehensive update of caspase substrates. *Cell Death Differ.* **10**, 76–100 (2003).
9. Elmore, S. Apoptosis: a review of programmed cell death. *Toxicol. Pathol.* **35**, 495–516 (2007).
10. Westphal, D., Dewson, G., Czabotar, P. E. & Kluck, R. M. Molecular biology of Bax and Bak activation and action. *Biochim. Biophys. Acta* **1813**, 521–531 (2011).

11. Tait, S. W. & Green, D. R. Mitochondria and cell death: outer membrane permeabilization and beyond. *Nat. Rev. Mol. Cell Biol.* **11**, 621–632 (2010).
12. Pradelli, L. A. et al. Glycolysis inhibition sensitizes tumor cells to death receptors-induced apoptosis by AMP kinase activation leading to Mcl-1 block in translation. *Oncogene* **29**, 1641–1652 (2010).
13. Alves, N. L. et al. The Noxa/Mcl-1 axis regulates susceptibility to apoptosis under glucose limitation in dividing T cells. *Immunity* **24**, 703–716 (2006).
14. Coloff, J. L. et al. Akt-dependent glucose metabolism promotes Mcl-1 synthesis to maintain cell survival and resistance to Bcl-2 inhibition. *Cancer Res.* **71**, 5204–5213 (2011).
15. Quirce, R. et al. New insight of functional molecular imaging into the atheroma biology: 18F-NaF and 18F-FDG in symptomatic and asymptomatic carotid plaques after recent CVA. Preliminary results. *Clin. Physiol. Funct. Imaging* **36**, 499–503 (2016).
16. Sharma, V., Ichikawa, M. & Freeze, H. H. Mannose metabolism: more than meets the eye. *Biochem. Biophys. Res. Commun.* **453**, 220–228 (2014).
17. Chattopadhyay, U. & Bhattacharyya, S. Inhibition by monosaccharides of tumor associated macrophages mediated antibody dependent cell cytotoxicity to autologous tumor cells. *Neoplasma* **34**, 295–303 (1987).
18. Gartner, S. L. & Williams, T. J. Modulation of interleukin-1 induced thymocyte proliferation by D-mannose. *Thymus* **19**, 117–126 (1992).
19. Alton, G. et al. Direct utilization of mannose for mammalian glycoprotein biosynthesis. *Glycobiology* **8**, 285–295 (1998).
20. De Robertis, M. et al. The AOM/DSS murine model for the study of colon carcinogenesis: from pathways to diagnosis and therapy studies. *J. Carcinog.* **10**, 9 (2011).

Author contributions K.M.R., P.S.G. and A.D.B. conceived the study. P.S.G., J.O., J.-i.S., F.B., E.K. and S.R. conducted and analysed cell growth, cell death and enzyme assays and western blotting. P.S.G., S.C. and G. Mackay conducted and analysed metabolic experiments. V.J.A.B., D.M.G. and B.Z. performed animal experiments. G. Malviya and A.M. performed and analysed PET and MRI scans. C.N. performed immunohistochemistry. A.R., D.E., A.H. and D.M. generated and analysed tissue microarrays. P.S.G. and K.M.R. wrote the manuscript. I.A.M., O.J.S., J.E. and K.M.R. supervised the study.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0729-3.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0729-3.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to K.M.R.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Cell culture, transfections and infections. All cell lines were from Beatson Institute stocks and were originally obtained from American Type Culture Collection or European Collection of Authenticated Cell Cultures repositories, apart from PATU-8902 (DMSZ, ACC-179) and KP-4 (RIKEN, RCB1005). A549, IGROV-1, Saos-2, U2OS, U2OS-E1a, SKOV-3, RKO and PATU-8902 cells were grown in DMEM high glucose supplemented with 10% FBS, glutamine (0.292 mg ml^{-1}) and penicillin (100 units per ml)–streptomycin (100 µg ml^{-1}) (all from Life Technologies). U2OS-E1a cells were generated in our laboratory and have been previously described[21]. K562 cells were grown in RPMI-1640 and KP-4 cells were grown in IMDM (Life Technologies) with 20% FBS, supplemented with penicillin (100 units per ml)–streptomycin (100 µg ml^{-1}). The B16-F1 mouse skin melanoma cell line was a gift from L. Machesky (Cancer Research UK Beatson Institute) and cells were maintained in DMEM (Thermo Fisher Scientific, 21969035) supplemented with 10% FBS (Thermo Fisher Scientific, 10270106), 2 mM L-glutamine (Thermo Fisher Scientific, 25030032), and penicillin–streptomycin (Thermo Fisher Scientific, 15140122). LLC mouse Lewis lung carcinoma cells were a gift from S. Zanivan (Cancer Research UK Beatson Institute) and were maintained in RPMI-1640 (Thermo Fisher Scientific, 31870074), 10% FBS, 10 mM HEPES (Thermo Fisher Scientific, 15630080), 2 mM L-glutamine and penicillin–streptomycin. These cell lines were confirmed to be free of mycoplasma.

Cells were transfected using calcium phosphate precipitates as previously described[16]. U2OS-E1a cells were infected with virus containing pBabe-puro-empty or pBabe-puro-Bcl-X$_L$ following the same protocol as previously published[21]. U2OS-E1a cells were infected with virus containing pLZRS empty or pLZRS HA-Mcl-1 (provided by S. Tait)[22]. U2OS-E1a-infected cells were selected in DMEM 10% FBS containing 600 µg ml^{-1} of neomycin for 2 weeks. U2OS control and Bax/Bak CRISPR cells were provided by S. Tait.

The following siRNAs were used: PMI siRNA-1 (Dharmacon, J-011729-05-0002), PMI siRNA-2 (Dharmacon, J-011729-06-0002), PMI siRNA-3 (Dharmacon, J-011729-07-0002), PMI siRNA-4 (Dharmacon, J-011729-08-0002), NTC1 (Dharmacon, D-001810-03-20) and NTC2 (Dharmacon, D-001810-04-20).

lentiCRISPR v2 was a gift from F. Zhang (Addgene plasmid 52961)[23]. The following single-guide RNA sequences were used in the experiments. NTC 1: GTAGCGAACGTGTCCGGCGT; NTC 2: GCTTGAGCACATACGCGAAT; ATG5: AAGAGTAAGTTATTTGACGT; ATG7: GAAGCTGAACGAGTATCGGC; Casp8: GCCTGGACTACATTCCGCAA; FADD: TTCCTATGCCT CGGGC GCGT; BAX: AGTAGAAAAGGGCGACAACC; BAK: GCCATGCTGGTAGAC GTGTA; NOXA: TCGAGTGTGCTACTCAACTC.

The following shRNAs were used: pGIPZ-non-targeting control (NTC) (Dharmacon, RHS4346), PMI 1 (Dharmacon, RMM4431-200352145, Clone ID: V2LMM_110673) and PMI 2 (Dharmacon, RMM4431-200355616, Clone ID: V2LMM_203337).

pLX304 was a gift from D. Root (Addgene plasmid 25890). Human PMI cDNA was amplified using pCMV-Sport-MPI (Dharmacon, MHS6278-202802339) as a template and inserted into pDONR221 vector (Thermo Fisher Scientific, 12536017) using Gateway BP Clonase II Enzyme mix (Thermo Fisher Scientific, 11789020). PMI cDNA was then transferred into a pLX304 destination vector using Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific, 11791020). Lentivirus production and infection were carried out as described previously[24].

Cell culture treatments. *Mannose treatment.* Cells were seeded in six-well plates and incubated overnight at 37 °C. The following day, the medium was replenished with fresh full growth medium (DMEM, RPMI-1640 or IMDM) containing 25 mM D-mannose (DMEM or IMDM) or 11.11 mM D-mannose (RPMI-1640). Other sugars were added to a concentration of 25 mM: D-glucose (Sigma-Aldrich, G8270), D-fructose (Sigma-Aldrich, F3510), D-galactose (Sigma-Aldrich, G5388), L-fucose (Sigma-Aldrich, F2252 and Cayman Chemical, 16479). New stocks were prepared every two weeks of 1 M mannose in Milli-Q water and sterilized by filtering through a 0.22-µm pore filter. For control conditions, the same volume of Milli-Q water was added to the medium. Cells were left for at least 24 h for cell-death experiments or for 6 h for metabolic experiments. Cell death was blocked by treatment with zVAD-FMK (Apax Labs, A1902). Where indicated, cells were also treated with 2-deoxy-D-glucose (Sigma-Aldrich, D8375), tunicamycin (Sigma-Aldrich, T7765) or chloroquine (Sigma-Aldrich, C6628).

13C-labelled sugars. The day after seeding, cells were washed three times with abundant PBS before adding glucose-free DMEM (supplemented with 10% dialysed FBS, 2 mM glutamine, 100 units per ml of penicillin and 100 µg ml^{-1} of streptomycin). Depending on the experiments, this medium could contain 5 mM of 1,2-^{13}C$_2$-glucose alone, together with ^{13}C$_6$-mannose or 5 mM ^{13}C$_6$-mannose alone. Stocks of 1,2-^{13}C$_2$-glucose and ^{13}C$_6$-mannose were prepared at a concentration of 0.5 M in Milli-Q water before sterilization by filtering through a 0.22-µm pore filter. *Chemotherapeutic drugs.* Cells were plated overnight and then incubated for the indicated times in control or mannose-containing medium. After one day of incubation, fresh control and mannose-containing media were prepared and drugs were added as described for each experiment. The drugs used were cisplatin (Sigma-Aldrich, C2210000) and doxorubicin (Sigma-Aldrich, D1515).

RT–qPCR. RT-qPCR was carried out as previously described[24] using primers for MCL1 (Qiagen, QT 00094122) and BCL2L1 (Qiagen, QT00997423). mRNA levels were determined by the relative standard curve method, normalized to 18S.

Western blotting. Protein extraction was performed as previously described[25]. Protein lysates were separated by SDS–PAGE and blotted onto nitrocellulose membranes. Western blot analysis was performed according to standard techniques as previously described[25]. The following antibodies were used at a dilution of 1:1,000 unless otherwise stated: anti-β-actin (Abcam, ab8227), anti-Mcl-1 (Cell Signaling, 4572), anti-ERK2 (Santa Cruz, sc-154), anti-Bcl-X$_L$ (Cell Signaling, 2762), anti-HSP-90β (Santa Cruz, sc-1057), anti-PARP (Cell Signaling, 9542), anti-cleaved caspase-3 (Cell Signaling, 9664), anti-FADD (BD Transduction Laboratories, F36620), anti-caspase-8 (Cell Signaling, 4790), anti-Bax (BD Transduction Laboratories, 610983), anti-Bak (Cell Signaling, 6947), anti-Bim (Cell Signaling, 2993), anti-Noxa (Novus Biologicals, NB-600-1159), anti-phospho-AMPKα (Cell Signaling, 2535), anti-AMPKα (Cell Signaling, 5832), anti-PMI (Abcam, 128115), anti-LC3B (Cell Signaling, 2775S, 1:1,500), anti-β-actin (Cell Signaling, 4670S, 1:2,000), anti-ATG5 (Cell Signaling, 12994S), anti-ATG7 (Cell Signaling, 8558S), anti-Bip (also known as GRP78; Cell Signaling, 3177S) and anti-p62 (BD Biosciences, 610833, 1:2,000). Mouse PMI protein was detected using rabbit polyclonal PMI antibody (Proteintech, 14234-1-AP). Validation was based on information provided in the manufacturers' datasheets. In addition, as indicated in the manuscript, we used RNAi or CRISPR–Cas9 to validate the antibodies used to detect the following proteins: Bax, Bak, caspase-8, FADD, PMI, Atg5 and Atg7.

Metabolic extraction of intracellular metabolites. Cells were seeded at a concentration of 100,000 cells per well in six-well plates. The next morning, the medium was replenished with fresh full medium and cells were kept under these conditions for another 24 h to stabilize their metabolism. Approximately 36–40 h after being plated, cells were treated under different conditions as described, using full growth medium or glucose-free medium in the presence or absence of unlabelled sugars or labelled sugars for 6 h. After a 6-h incubation at 37 °C, intracellular metabolites were extracted.

The medium was aspirated and six-well plates were placed on ice and washed thoroughly with 4 °C PBS three times before the addition of 500 µl extraction solvent (50% methanol, 30% acetonitrile, 20% Milli-Q water) to each well. Plates were then agitated at 4 °C for 5 min to successfully extract intracellular metabolites and then centrifuged at 16,000g for 10 min at 4 °C. Supernatants were transferred into HPLC vials and stored at −80 °C before LC–MS analysis.

An Exactive Orbitrap mass spectrometer (Thermo Scientific) was used together with a Thermo Scientific Accela HPLC system. The HPLC setup consisted of a ZIC-pHILIC column (SeQuant, 150 mm × 2.1 mm, 5 µm, Merck KGaA) with a ZIC-pHILIC guard column (SeQuant, 20 mm × 2.1 mm) and an initial mobile phase of 20% 20 mM ammonium carbonate, pH 9.4 and 80% acetonitrile. Cell extracts (5 µl) were injected and metabolites were separated over a 15-min mobile phase gradient, decreasing the acetonitrile content to 20%, at a flow rate of 200 µl min^{-1} and a column temperature of 45 °C. The total analysis time was 23 min. For longer runs of 37 min, a 30-min gradient with the same solvents was used, at a flow rate of 100 µl min^{-1} and a column temperature of 30 °C[26]. All metabolites were detected across a mass range of 75–1,000 m/z using the Exactive mass spectrometer at a resolution of 25,000 (at 200 m/z), with electrospray ionization and polarity switching to enable both positive and negative ions to be determined in the same run. Lock masses were used and the mass accuracy obtained for all metabolites was below 5 p.p.m. Data were acquired with Thermo Xcalibur software (version 2.2).

The peak areas of different metabolites were determined using Thermo TraceFinder software (version 3.2), in which metabolites were identified by the exact mass of the singly charged ion and by known retention time on the HPLC column. Commercial standards of all metabolites detected had been analysed previously on this LC–MS system with the pHILIC column. The ^{13}C labelling patterns were determined by measuring peak areas for the accurate mass of each isotopologue of many metabolites. Intracellular metabolites were normalized to the protein content of the cells, measured at the end of the experiment by the Lowry assay[27]. As the proteins precipitate upon addition of the metabolite extraction solvent, protein content was measured in the wells after the metabolites were extracted.

Translation assay. Cells at 50% confluency were pretreated with or without 25 mM mannose for 24 h or with 100 µg ml^{-1} cycloheximide for 1 h. [^{35}S]methionine (1 MBq) (Perkin Elmer, EasyTag EXPRESS^{35}S Protein Labelling Mix, NEG772002MC) was added to the culture medium (2 ml per well in a six-well plate) for 30 min. Cells were washed in ice-cold PBS, then lysed in lysis buffer (10 mM Tris pH 7.5, 50 mM NaCl, 0.5% NP40, 0.5% SDS, benzonase (Sigma-Aldrich E1014, 2 µl per 10 ml of lysis buffer)). Proteins were precipitated in 25% trichloroacetic acid at 4 °C for 30 min. The precipitates were washed on glass-fibre filters (Whatman 934-AH, 1287-024) with 70% ethanol followed by acetone, dried, and

incorporation of ^{35}S was quantified in a liquid scintillation counter. The results were calculated as relative ^{35}S incorporation per 10^5 cells.

Transcription assay. Cells at 50% confluency were pretreated with 25 mM mannose or with a control for 24 h, or with 5 μM actinomycin D for 1 h. [^{32}P]UTP (1.1 MBq) (Perkin Elmer, BLU007H001MC) was added to the culture medium (2 ml per well in a six-well plate) for 6 h. Cells were washed in PBS and mRNA was prepared using the Dynabeads mRNA DIRECT Kit (Life Technologies, 61012). ^{32}P-labelled mRNA was quantified in a liquid scintillation counter and the results were calculated as relative [^{32}P]UTP incorporation per 10^5 cells.

PMI enzymatic assay. Cells were grown to confluence, washed with PBS and collected by centrifugation (5 min, 150g at 4 °C). Cell pellets were lysed by three freeze–thaw cycles. Post-nuclear protein fractions (40 μg) were used to determine PMI activities in each cell line tested by means of a coupled enzymatic reaction. In brief, samples were incubated in a buffer containing 200 mU of PGI (Roche, 10 127 396 001), 500 mM glucose-6-phosphate dehydrogenase, 1 mM NADP$^+$, 40 mM Tris-HCl pH 7.4, 6 mM MgCl$_2$, 5 mM Na$_2$HPO$_4$/KH$_2$PO$_4$. Reactions were initiated by the addition of 1 mM mannose-6-phosphate, and the production of NADPH/H$^+$ was assessed for 2 h at room temperature by measuring the optical density at 340 nm (OD$_{340nm}$). In parallel, western blots directed against PMI and ERK2 were performed to examine the correlation between PMI activities and PMI expression levels in each cell line analysed.

Proteasome assay. Cells were seeded 24 h before 24-h treatments using media either with or without mannose supplementation (25 mM). A luminescence-based assay (Proteasome-Glo chymotrypsin-like cell-based assays, Promega) was performed according to the manufacturer's protocol. The measured proteasome activities were normalized to cell number by counting cells at the end of each experiment. Cells were treated with 10 μM MG132 as a control for the specificity of the assay.

Lowry assay. The Lowry assay was used to determine the total protein content for each one of the triplicates from the metabolic experiments. First, 500 μl of solution A (70% Milli-Q water, 20% 5 M NaOH, 10% 2,5-dimethoxy-4-chloroamphetamine) was added to each well from a six-well plate and left under agitation for 20 min. Next, 5 ml of solution B (0.5 g NaCu-EDTA, 40 g Na$_2$CO$_3$, 8 g NaOH in 2 l) was added to each well and left for 40 min. Then, 500 μl of Folin reagent was added to each well and left for a minimum of 15 min until a blue colour was observed in each well. One or two six-well plates were used to construct a standard curve of protein concentration with BSA. Finally, 200 μl from each well was separately transferred to a 96-well plate and the absorbance was measured at 750 nm. The protein concentration was then measured by calculating the equation of the standard curve and extrapolating the absorbance of each well.

Flow cytometry. Flow cytometry for unfixed cells (cell-death assay) was conducted as previously described[28], and a FACSCalibur or ATTune NXT flow cytometer was used for the analysis.

Animal experiments. All in vivo xenograft experiments were performed using six-week old female CD1-nude mice or C57/BL6J wild-type mice as approved by the Glasgow University Animal Welfare and Ethical Review Body and in accordance with UK Home Office guidelines. Mice were placed five per cage with free access to water and food (chow diet). Experimental cohort sizes were based on previous similar studies that have given statistical results while also respecting the limited use of animals in line with the 3R system: replacement, reduction, refinement. In no cases were mice maintained once the tumour size limit that was permitted in our Home Office license (15 mm in any direction) had been reached. All treatment studies were randomized, but did not involve blinding.

KP-4 cells (5×10^6 cells) were injected subcutaneously with 100 μl of Matrigel in either the right flank or both the left and the right flanks. LLC cells (1×10^6 cells) and B16-F1 cells (1×10^6 cells) were injected subcutaneously in 200 μl or 100 μl of PBS suspension, respectively. Mice were weighed and tumours were measured using callipers three times per week (usually on Monday, Wednesday and Friday of each week). The calculation of tumour volume was as follows: $(L \times S^2)/2$ (where L is the longest length and S is the shortest length).

For mannose treatment, normal drinking water was exchanged for 200 ml of 20% mannose in drinking water (w/v) and it was replaced once every week. Mice received mannose by oral gavage (200 μl) three times per week from the same stock of 20% mannose in water.

For doxorubicin treatments, each mouse received intraperitoneal injections at a concentration of 5 mg kg^{-1} once per week. Stocks of 1 mg ml^{-1} doxorubicin (Sigma-Aldrich, D1515) were prepared in Milli-Q water.

For azoxymethane (AOM) + dextran sodium sulfate (DSS) experiments, mice were injected with a single dose of AOM (Sigma-Aldrich, A5486) at 10 mg kg^{-1}. Five days after AOM injection, mice received three cycles of 1.5% DSS (MP Biomedical, 0216011080) in water; each cycle lasted five consecutive days, with one week in between each cycle. In the week between each cycle, mice received normal water or 20% (w/v) mannose water. After the last cycle of DSS, mice were maintained with normal water or mannose for two weeks before being killed on day 68.

Tissue collection. All mice were killed by CO$_2$-mediated euthanasia[29] before collecting any tissue sample. Mice were placed in a cage and exposed to CO$_2$ for 100 s before neck dislocation.

Xenografted tumours were removed and detached from the skin and cut in half. One half of each tumour was fixed for 24–30 h in 10% neutral buffered formalin at room temperature. Formalin was then exchanged for 70% ethanol before histology.

The colons of mice from AOM + DSS carcinogenesis experiments were collected and fixed in formalin for histological processing.

For the experiments with the genetically engineered mouse model, *Villin*^{creER}*Apc*$^{fl/+}$*Kras*$^{G12D/+}$ mice[30] aged 6–12 weeks were given a single intraperitoneal injection of 80 mg kg^{-1} tamoxifen (Sigma-Aldrich, T5648). Four days post-induction, drinking water was exchanged for fresh drinking water or 20% mannose in drinking water (w/v) and given freely, this was replaced every week. Mice were aged until the clinical end point—when they displayed anaemia, hunching or weight loss. Colons were pinned out into 10% neutral buffered formalin and the number of tumours was counted.

Blood sampling and blood metabolic extraction. To measure blood mannose levels, tail tipping was performed. Blood samples were directly frozen using dry ice and stored at −80 °C until metabolic extraction was performed.

For metabolic extraction of whole blood, 2–5 μl of each sample was diluted in 100–250 μl of metabolic extraction buffer on ice for 5 min (1:50 dilution). Then, all samples were centrifuged at 17,949g at 4 °C for 15 min. Finally, 100 μl of the supernatant from each sample was transferred into HPLC vials and stored at −80 °C until LC–MS analysis.

PET and MRI scanning. Mice with KP-4 cells xenografts ($n = 9$, weight 24.6 ± 1.8 g) received either 200 μl of 20% w/v mannose in water (treatment group) or normal water (control group) by oral gavage 20 min before the injection of [^{18}F]FDG. [^{18}F]FDG (12.9 ± 1MBq) in 200 μl of normal saline was administered via an intravenous bolus injection in the tail vein. After an uptake phase of 30 min, PET and MRI scans were sequentially performed using a nanoScan (Mediso Ltd) PET/MRI (1T) scanner. Mice were maintained under 2–2.5% isoflurane in medical air during the injection and imaging procedures. Static PET acquisitions were performed for 15 min, and subsequently whole body T1 GRE 3D Multi-FOV MRI scans (slice thickness 0.50 mm, repetition time (TR) 10 ms, echo time (TE) 2.3 ms, flip angle 12°) were performed to obtain anatomical references. For quantitative assessments of scans, regions of interest (ROI) were manually drawn around the edge of the tumour xenograft on MRI scans by visual inspection using PMOD software version 3.504 (PMOD Technologies), and the same ROI was copied on the respective PET scans. Tumour ROIs were slightly different between scans depending on the positions and angles of the mice on the scanner, therefore separate ROIs for each scan were drawn. The percentage injected dose per ml (%ID ml^{-1}) was calculated using the formula %ID ml^{-1} = ROI activity (kBq ml^{-1}) / injected dose × 100%. Data were reported as average %ID ml^{-1} ± s.d. Student's t-tests were used when comparing data between mannose-treated and control mice.

Tissue microarray PMI staining. Tissue microarrays (TMAs) were first placed in xylene for 5 min before three washes of 1 min each (two in ethanol and one in 70% ethanol). TMAs were washed for 5 min in deionized water and then held for 25 min at 98 °C in a PT module using pH 6 sodium citrate retrieval buffer. They were further washed once in tris-buffered Tween (TbT) before blocking endogenous peroxidase for 5 min. TMAs were washed again using TbT and stained with PMI antibody (1C7, 1:50 dilution) for 1 h. TMAs were washed in TbT before and after incubation for 30 min in a mouse EnVision detection system. A 10-min incubation in 3,3′-diaminobenzidine tetrahydrochloride was then perfomed. TMAs were again washed with deionized water (1 min) before and after incubation with haematoxylin Z, followed by post-incubation with 1% acid alcohol. This was followed by a 30-s wash with deionized water, and then incubation with Scott's Tap Water substitute (1 min) and deionized water (1 min).

BrdU staining. Xenografted tumours were embedded in paraffin, after which sections of an appropriate thickness were cut and kept at 60 °C overnight. Sections were first placed in xylene for 5 min before three washes of 1 min each (two in ethanol and one in 70% ethanol). Sections were washed for 5 min in deionized water and then held for 25 min at 98 °C in a PT module using pH 6 sodium citrate retrieval buffer. They were further washed once in TbT before blocking endogenous peroxidase for 5 min. Sections were washed in TbT before and after incubation with BrdU antibody (BD Biosciences 347580, 1:200 dilution) for 35 min. Sections were washed in TbT before and after incubation for 30 min in a mouse EnVision detection system. A 10-min incubation in 3,3′-diaminobenzidine tetrahydrochloride was then perfomed. Sections were again washed with deionized water (1 min) before and after incubation with haematoxylin Z, followed by post-incubation with 1% acid alcohol. This was followed by a 30-s wash with deionized water, and then incubation with Scott's Tap Water substitute (1 min) and deionized water (1 min).

Tissue microarrays. TMAs from ovarian, breast, colorectal, prostate and renal cancer were stained for PMI. The microarrays were scanned using a digital slide scanner (SCN 400 F, Leica Biosystems) and were scored for PMI expression on the basis of the percentage of tumour that showed negative, low, medium or high expression. Each sample was scored with a number between 0 and 300 according to the following equation: Score = (0 × negative) + (1 × low) + (2 × medium) + (3 × high).

Statistical analysis and reproducibility. All presented data were analysed using GraphPad Prism software. Four different statistical analyses were performed depending on the data from the different experiments shown, typically one-way ANOVA, two-way ANOVA, log-rank (Mantel–Cox) test and Student's t-tests that could be paired or unpaired and one-tailed or two-tailed. Four levels of significance were determined: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$, with NS indicating no significance. Where a representative experiment is shown, the number of times the same result was observed in independent experiments is detailed in the corresponding figure legend.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability
The data supporting the findings of this study are available within the paper and its Supplementary Information. Source Data for Figs. 1–4 and Extended Data Figs. 1–10 are available with the online version of the paper. Data are available from the corresponding author upon reasonable request.

21. Long, J. S. et al. Extracellular adenosine sensing—a metabolic cell death priming mechanism downstream of p53. *Mol. Cell* **50**, 394–406 (2013).
22. Riley, J. S. et al. Mitochondrial inner membrane permeabilisation enables mtDNA release during apoptosis. *EMBO J.* **37**, e99238 (2018).
23. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
24. Sakamaki, J. I. et al. Bromodomain protein BRD4 is a transcriptional repressor of autophagy and lysosomal function. *Mol. Cell* **66**, 517–532 (2017).
25. Mrschtik, M. et al. DRAM-3 modulates autophagy and promotes cell survival in the absence of glucose. *Cell Death Differ.* **22**, 1714–1726 (2015).
26. Mackay, G. M., Zeng, L., van den Broek, N. J. F. & Gottlieb, E. Analysis of cell metabolism using LC–MS and isotope tracers. *Methods Enzymol.* **561**, 171–196 (2015).
27. Karvela, M. et al. ATG7 regulates energy metabolism, differentiation and survival of Philadelphia-chromosome-positive cells. *Autophagy* **12**, 936–948 (2016).
28. Rosenfeldt, M. T. et al. E2F1 drives chemotherapeutic drug resistance via ABCG2. *Oncogene* **33**, 4164–4172 (2014).
29. Rosenfeldt, M. T. et al. p53 status determines the role of autophagy in pancreatic tumour development. *Nature* **504**, 296–300 (2013).
30. Cammareri, P. et al. TGFβ pathway limits dedifferentiation following WNT and MAPK pathway activation to suppress intestinal tumourigenesis. *Cell Death Differ.* **24**, 1681–1693 (2017).
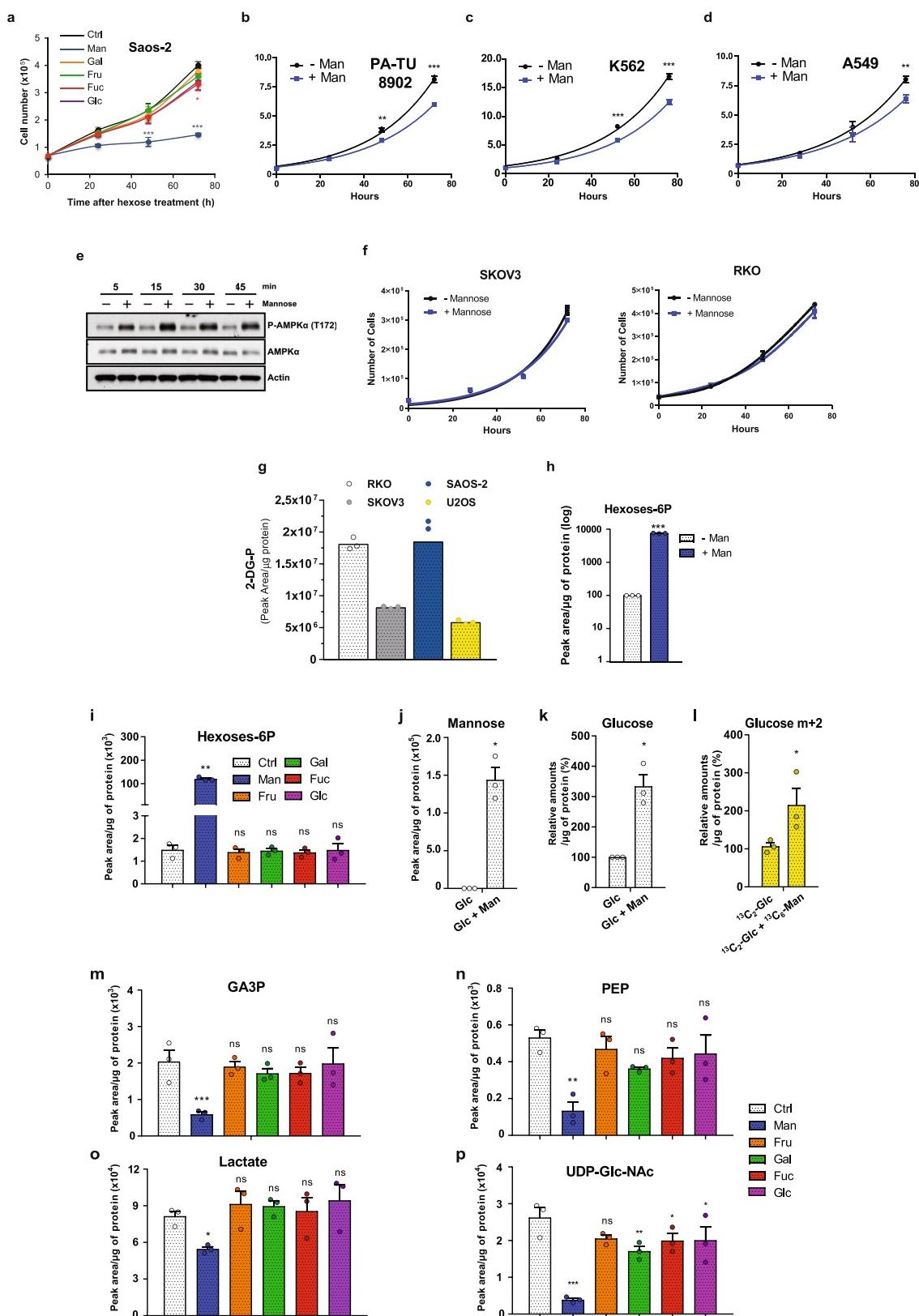
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Mannose affects cell growth and metabolism.
a, Growth curves of Saos-2 cells supplemented without (control) or with 25 mM of hexoses (man, mannose; gal, galactose; fru, fructose; fuc, fucose; glc, glucose). **b–d**, Growth curves of PA-TU-8902 (**b**) and A549 cells (**d**) in DMEM alone or with an additional 25 mM mannose; K562 cells in 10% FBS RPMI-1640 medium with or without 11.1 mM mannose (**c**).
e, Western blots showing the levels of phospho-AMPKα (T172) and total AMPKα after 5, 15, 30 and 45-min incubation of U2OS-E1a with standard medium or medium supplemented with 25 mM mannose. **f**, Growth curves of SKOV3 and RKO cells in DMEM alone or with an additional 25 mM mannose. **g**, Levels (expressed as peak area per microgram of protein) of 2-deoxyglucose-phosphate (2-DG-P) in RKO, SKOV3, SAOS-2 and U2OS-E1a (U2OS) cells incubated with 10 mM 2-deoxyglucose for 6 h in the presence of 25 mM mannose in the culture medium (DMEM). Data are the average of three technical replicates and are representative of two independent experiments. **h**, Levels (measured as the percentage of peak area per microgram of protein, on a $\log_{10}$ scale) of hexoses-6-phosphate (hexoses-6P) in U2OS-E1a cells after 6 h incubation in 10% dialysed FBS with 5 mM glucose in DMEM either with or without 5 mM

mannose. **i**, Peak area per microgram of protein of hexoses-6-phosphate in U2OS-E1a cells incubated for 6 h in DMEM, with or without an additional 25 mM of the indicated sugars. **j**, **k**, Peak area per microgram of protein of intracellular non-phosphorylated mannose (**j**) or relative amounts of glucose (**k**) after a 6 h incubation of U2OS-E1a cells in 5 mM glucose in DMEM, with or without 5 mM mannose. **l**, Relative amount per microgram of protein of non-phosphorylated glucose $m + 2$ after 6 h incubation of U2OS-E1a cells in glucose-free DMEM either with 5 mM $1,2$-$^{13}C_2$-D-glucose alone or with 5 mM $1,2$-$^{13}C_2$-D-glucose and 5 mM $^{13}C_6$-D-mannose. **m–p**, Peak area per microgram of protein of glyceraldehyde-3-phosphate (GA3P) (**m**), phosphoenolpyruvate (PEP) (**n**), lactate (**o**) and UDP-GlcNAc (**p**) in U2OS-E1a cells after 6 h incubation in DMEM, with or without an additional 25 mM of the indicated sugars. $n = 3$ independent experiments (**a–d**, **f**, **h–p**). Data are representative of two independent experiments (**e**). All data are mean ± s.e.m. and were analysed by one-way ANOVA (**i**, **m–p**), two-way ANOVA followed by Tukey's multiple comparisons (**a**) or paired two-tailed Student's t-test (**h**, **j–l**). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
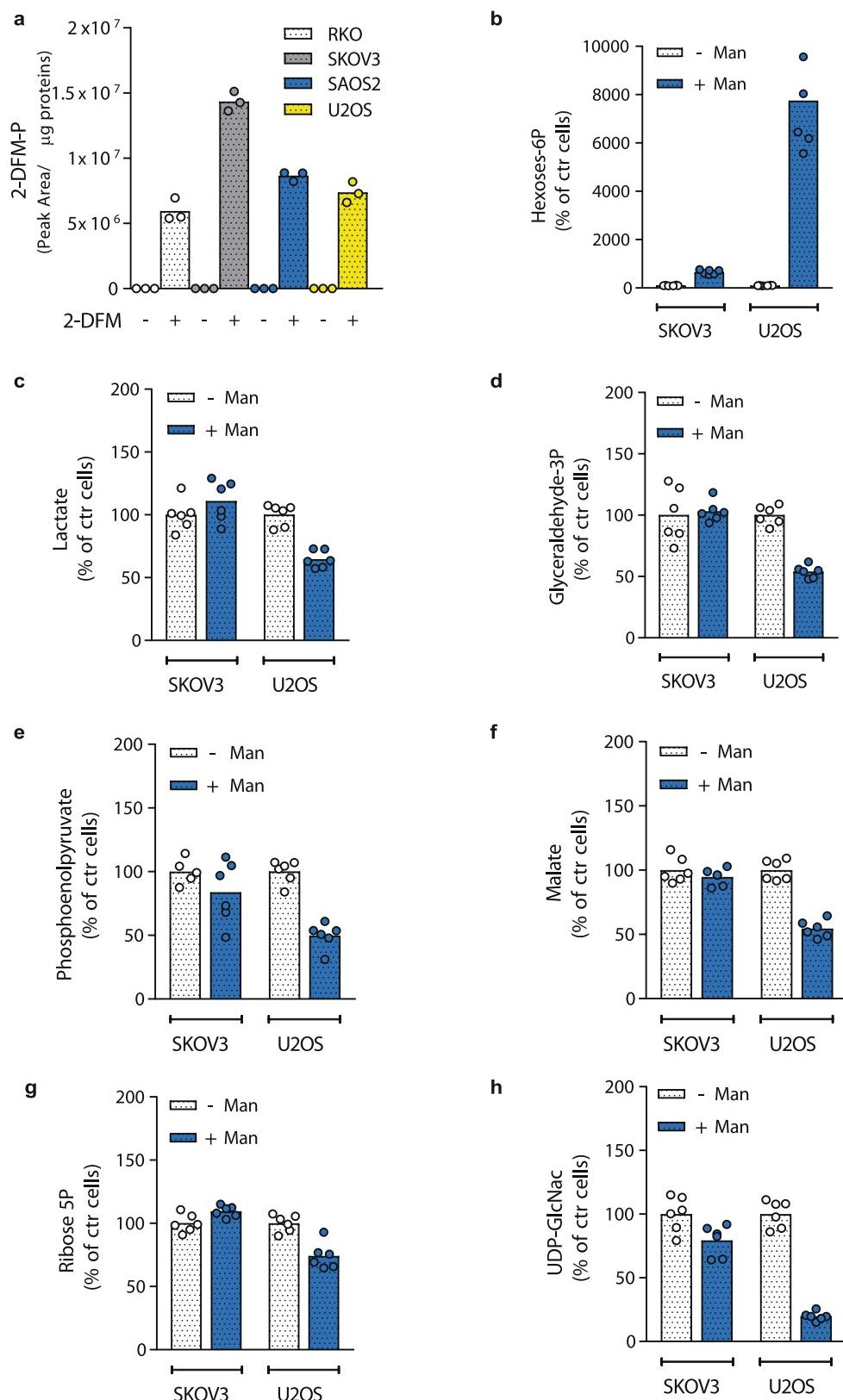
Extended Data Fig. 2 | Mannose sensitivity is not associated with mannose uptake but with changes in cellular metabolism. a, Levels (expressed as peak area per microgram of proteins) of 2-deoxy-2-fluoro-D-mannose-phosphate (2-DFM-P) in RKO, SKOV3, SAOS-2 and U2OS-E1a (U2OS) cells after 6 h treatment with or without 2 mM 2-deoxy-2 fluoro-D-mannose. **b–h**, Relative levels of hexoses-6-phosphate (**b**), lactate (**c**), glyceraldehyde-3-phosphate (glyceraldehyde-3P)

(**d**), phosphoenolpyruvate (**e**), malate (**f**), ribose-5-phosphate (ribose-5P) (**g**) and UDP-GlcNac (**h**) in SKOV3 and U2OS-E1a (U2OS) cells after 6 h treatment with or without 25 mM mannose in the culture medium (DMEM). Data in **a** are the average of three technical replicates and are representative of two independent experiments. In **b–h**, $n = 2$ independent experiments, each involving technical triplicates. All data are mean.
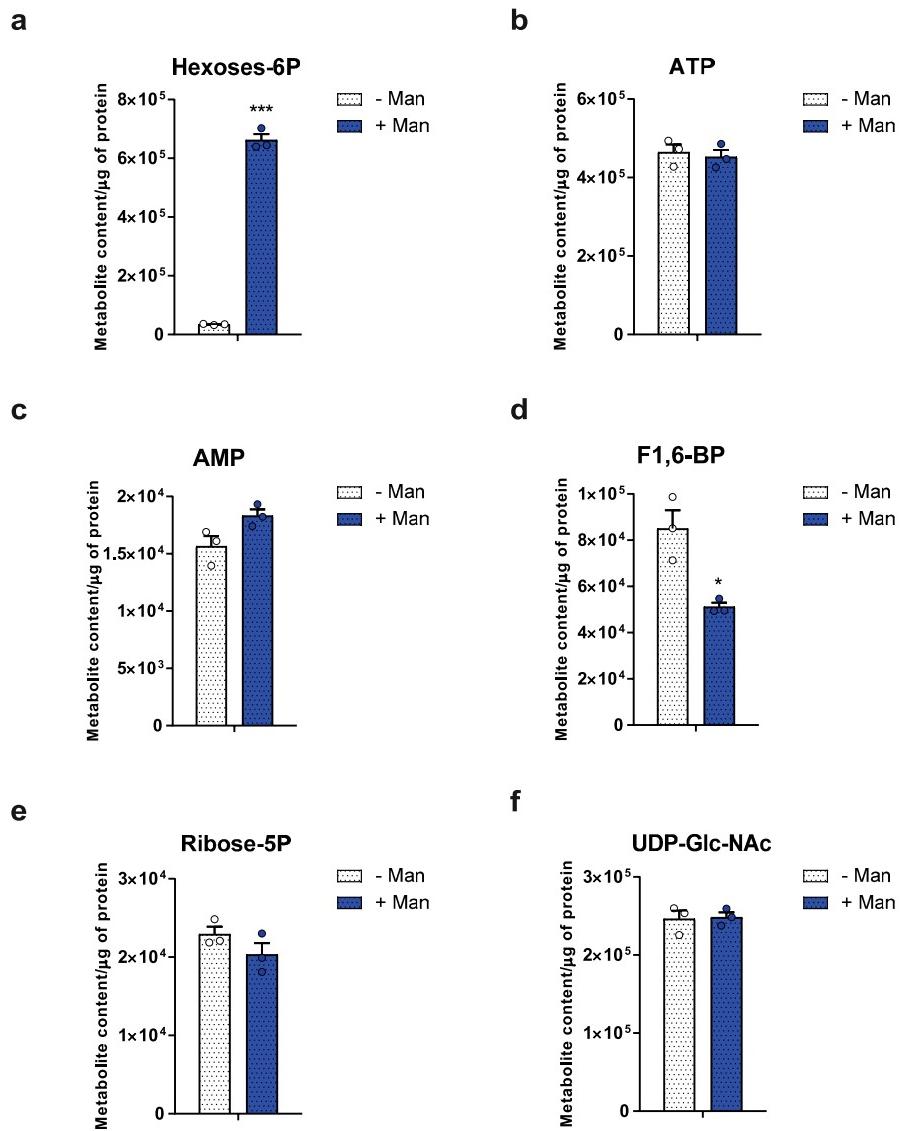
Extended Data Fig. 3 | Mannose has a rapid effect on cellular metabolism. Metabolite content (expressed as peak area per microgram of proteins) of hexoses-6-phosphate (**a**), ATP (**b**), AMP (**c**), fructose-1, 6-bisphosphate (F1,6-BP) (**d**), ribose-5-phosphate (**e**) and UDP-GlcNac (**f**) in U2OS-E1a (U2OS) cells after 5 min treatment with or without 25 mM mannose in the culture medium (DMEM). $n = 3$ independent experiments, presented as mean $\pm$ s.e.m. and analysed by a two-tailed unpaired t-test. *$P < 0.05$, ***$P < 0.001$.
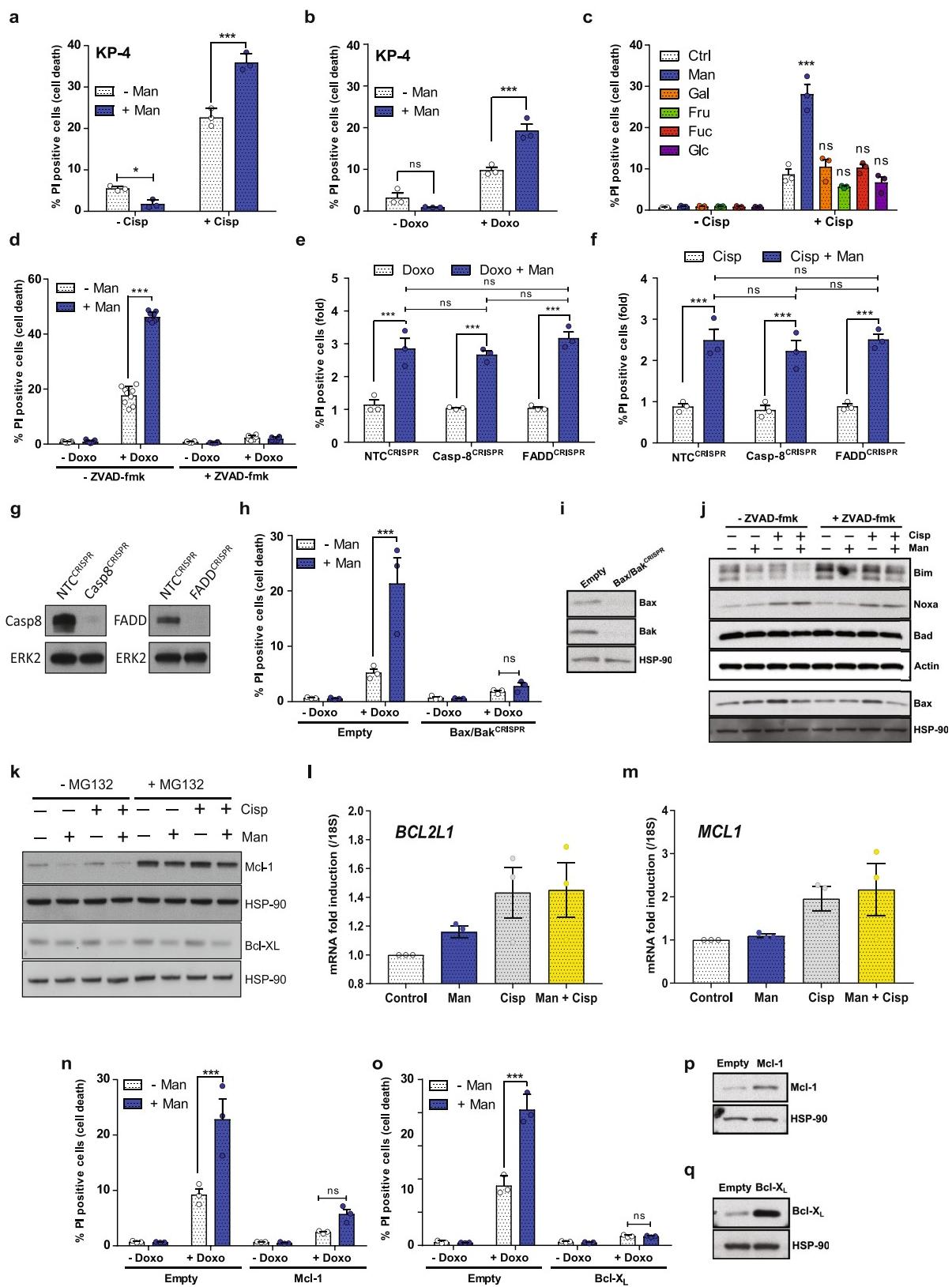
**Extended Data Fig. 4 |** See next page for caption.

Extended Data Fig. 4 | Mannose sensitizes cells to chemotherapy-induced cell death via the intrinsic apoptotic pathway. a, **b**, Percentage of propidium-iodide-positive KP-4 cells after 24 h treatment with 40 μM cisplatin (**a**) or 1 μg ml^{-1} doxorubicin (**b**) in the presence or absence of 25 mM mannose. **c**, Percentage of U2OS-E1a propidium-iodide-positive cells after 24 h treatment with or without 10 μM cisplatin in DMEM with or without 25 mM of the indicated additional sugars. **d**, Percentage of U2OS-E1a propidium-iodide-positive cells after 24 h treatment with or without 1 μg ml^{-1} doxorubicin, with or without 25 mM mannose and with or without 50 μM zVAD-FMK. **e**, **f**, Fold increase of the percentage of propidium-iodide-positive Saos-2 (NTC, caspase-8 and FADD CRISPR) cells upon 24 h treatment in DMEM with or without 1 μg ml^{-1} doxorubicin (**e**) or 10 μM cisplatin (**f**) and with or without 25 mM mannose. **g**, Western blots showing the levels of caspase-8, FADD and ERK2 in NTC, caspase-8 and FADD CRISPR Saos-2 cells. **h**, Percentage of empty and Bax/Bak CRISPR U2OS-E1a propidium-iodide-positive cells with or without 1 μg ml^{-1} doxorubicin and with or without 25 mM mannose. **i**, Western blots showing the levels of Bax, Bak and HSP90 in empty and Bax/Bak CRISPR U2OS-E1a cells. **j**, Western blots showing the levels of Bim, Noxa, Bad, Bax, Actin and HSP-90 in U2OS-E1a cells after 24 h treatment with or without 10 μM cisplatin, with or without 25 mM mannose and with or without 50 μM zVAD-FMK. The HSP-90 blot is identical to the one shown in Fig. 2f as the blots are from the same experiment. **k**, Western blots showing the levels of Mcl-1 and Bcl-X$_L$ in U2OS-E1a cells after 48 h with or without 10 μM cisplatin, with or without 25 mM mannose and in the absence or presence of 10 μM MG132 as indicated. **l**, **m**, PCR with reverse transcription (RT–PCR) showing the levels of *BCL2L1* (Bcl-X$_L$) (**l**) and *MCL1* (**m**) mRNAs in U2OS-E1a cells after 48 h treatment with 10 μM cisplatin alone, 25 mM mannose alone, or both 10 μM cisplatin and 25 mM mannose, in the presence of 50 μM zVAD-FMK. **n**, **o**, Percentage of U2OS-E1a (empty), Mcl-1 and Bcl-X$_L$ overexpressing propidium-iodide-positive cells after 24 h treatment with or without 1 μg ml^{-1} doxorubicin and with or without 25 mM mannose. **p**, **q**, Western blot showing the levels of Mcl-1, Bcl-X$_L$ and HSP-90 in U2OS-E1a (empty), Mcl-1 and Bcl-X$_L$ overexpressing cells. $n = 3$ independent experiments (**a–c**, **e**, **f**, **h**, **l–o**). Data are representative of two independent experiments (**g**, **j**, **k**) or one experiment (**i**, **p**, **q**). In **d**, $n = 3$ independent experiments (each involving technical triplicates) (−zVAD-FMK) and $n = 2$ independent experiments (each involving technical triplicates) (+zVAD-FMK). All data are mean ± s.e.m. and were analysed by two-way ANOVA with Bonferroni correction (**a–f**, **h**, **n**, **o**). *$P < 0.05$, ***$P < 0.001$.
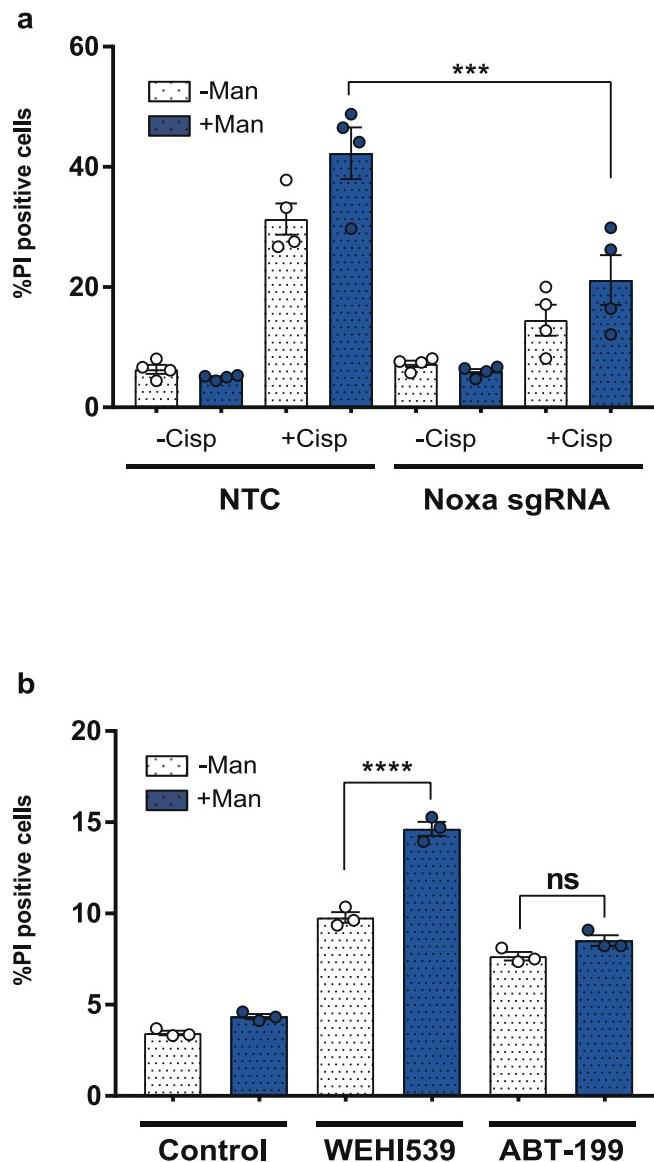
Extended Data Fig. 5 | Sensitization to cell death by mannose is connected to Bcl-2 family members. a, Percentage of control (NTC) and *Noxa* CRISPR U2OS-E1a propidium-iodide-positive cells upon 24 h treatment with or without 10 μM cisplatin and with or without 25 mM mannose. **b**, Percentage of U2OS-E1a propidium-iodide-positive cells upon 24 h treatment with 10 μM WEHI539 (a BclX$_L$ inhibitor) or 10 μM ABT-199 (a Bcl-2 inhibitor), with or without 25 mM mannose. $n = 4$ independent experiments (**a**); $n = 3$ three independent experiments (**b**). All data are mean ± s.e.m. and were analysed by two-way ANOVA with Bonferroni correction (**a**) and two-tailed unpaired *t*-test (**b**). ***$P < 0.001$; ****$P < 0.0001$; NS, not significant.
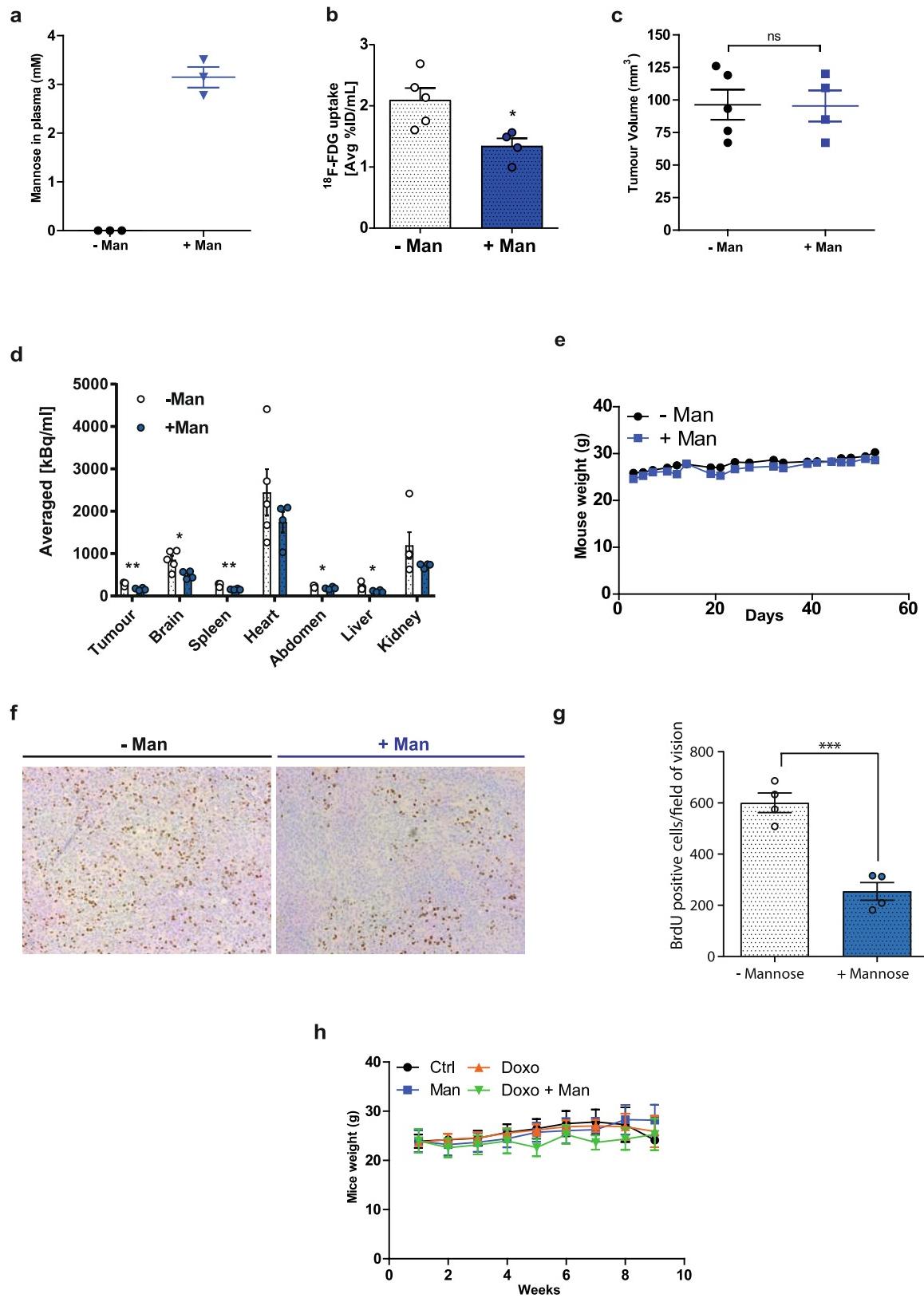
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Mannose affects cell proliferation and the uptake/retention of ^{18}F-FDG, but it does not affect animal weight.
a, Mannose levels in the plasma after 60 min in mice treated with a single dose of 200 μl water or 20% mannose in water. **b, c**, CD1-nude mice were transplanted with KP-4 cells subcutaneously and tumours were grown for 14 days. PET and MRI scans were performed for mice treated with 200 μl of water or 20% mannose in water by oral gavage 20 min before injection of [^{18}F]FDG into the tail vein. **b**, Quantification of [^{18}F]FDG uptake by tumours represented in average percentage injected dose per ml (%ID ml^{-1}) $\pm$ s.d. **c**, Volume of each tumour at the time of the PET and MRI scans. Data were analysed by unpaired two-tailed Student's t-test. **d**, CD1-nude mice were injected with KP-4 cells subcutaneously and treated with normal drinking water or 20% mannose in the drinking water, plus the same treatment (either normal water or 20% mannose) by oral gavage three days a week from the third day after tumour transplantation. Shown is the quantification of [^{18}F]FDG uptake by the tumour and different organs represented in averaged kBq ml^{-1} $\pm$ s.d. Data were analysed by unpaired two-tailed Student's t-test. **e**, CD1-nude mice were injected with KP-4 cells subcutaneously and treated with normal drinking water or 20% mannose in the drinking water, plus the same treatment (either normal water or 20% mannose) by oral gavage three days a week from the third day after tumour transplantation. The weight of mice was recorded at the indicated times. **f, g**, CD1-nude mice were injected with KP-4 cells subcutaneously and treated with normal drinking water or 20% mannose in the drinking water (either normal water or 20% mannose) by oral gavage three days a week from the third day after tumour transplantation. **f**, Images of BrdU sections representing tumours in control (left) and mannose-treated (right) mice. **g**, Quantification of BrdU-positive cells per section in control tumours ($n = 4$) and mannose-treated ($n = 4$) tumours. Five representative images per tumour were analysed. **h**, CD1-nude mice were injected with KP-4 cells subcutaneously and tumours were grown for 10 days before treatment with mannose and/ or doxorubicin (doxo) was started. Mice received normal drinking water (ctrl and doxo) or 20% mannose in the drinking water (man and doxo + man) together with the same treatment by oral gavage three times per week. Doxorubicin treatment started on day 32 and mice received 5 mg kg^{-1} by intraperitoneal injection once per week. The weight of mice in all groups was recorded throughout the experiment. The number of mice for each experiment is as follows: $n = 3$ per group (**a**), $n = 5$ (−mannose), $n = 4$ (+mannose) (**b**–**d**); $n = 10$ (**h**). In **a**, **c** and **h** data are mean $\pm$ s.e.m. Data were analysed with a two-tailed unpaired t-test (**c**). *$P < 0.05$, **$P < 0.01$.
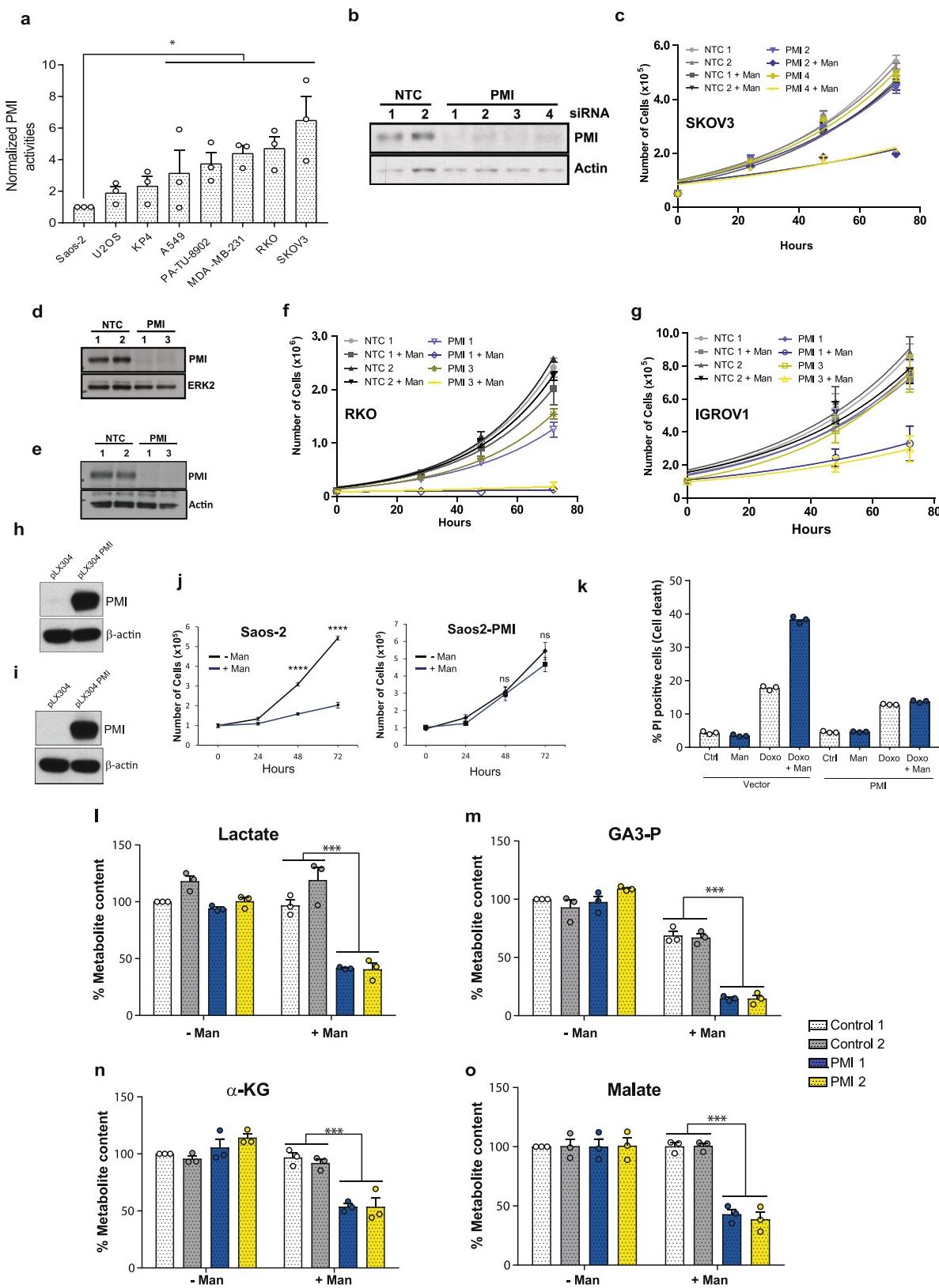
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | PMI levels dictate the response to mannose.
a, PMI expression levels correlate with PMI activities. PMI activities were
measured in eight different cell lines using coupled enzymatic reactions.
Graph shows the OD_{340nm} measured at 2 h, reflecting the amount of
$NADPH/H^+$ produced by the reactions. Results from three independent
experiments were normalized relative to PMI activities measured in
Saos-2 cells and represent mean ± s.e.m. **b–g**, *MPI* knockdown sensitizes
cells to mannose. **b**, Western blot of SKOV3-transfected cells showing the
levels of PMI and actin after 48 h of transient transfection with siRNAs.
Growth curves of SKOV3 (**c**), RKO (**f**) and IGROV1 (**g**) in regular DMEM
supplemented with or without 25 mM mannose after transient transfection
for 48 h with two NTC and two *MPI*-targeting siRNAs. **d, e**, Western
blot showing the levels of PMI and ERK2 in RKO (**d**) and PMI and actin
in IGROV1 (**e**) 48 h after siRNA transfection. **h–k**, Overexpression of
PMI causes resistance to the growth-suppressing and death-promoting
effects of mannose. PMI expression in U2OS-E1a (**h**) and Saos-2 (**i**) cells

was confirmed by western blotting. **j**, Saos2-PMI cells and control cells
(Saos-2) were plated in the presence or the absence of 25 mM mannose
and cell numbers were counted at the indicated times. **k**, Percentage of
propidium-iodide-positive U2OS-E1a control cells (vector) or U2OS-E1a
cells overexpressing PMI (PMI) after 24 h treatment with or without
$1\,\mu g\,ml^{-1}$ doxorubicin in the presence or absence of 25 mM mannose.
l–o, Percentage of lactate (**l**), GA3-P (**m**), α-KG (**n**) and malate (**o**)
metabolite content (peak area) in SKOV3 cells transfected with siRNA for
48 h before 6 h incubation in complete DMEM medium with or without
supplementation of 25 mM mannose. $n = 3$ independent experiments
(**a, c, f, g, j, l–o**). Data are representative of two independent experiments
(**b, d, e, k**), or one experiment (**h, i**). Data are mean ± s.e.m. and were
analysed by two-tailed unpaired t-test (**a**), two-way ANOVA with Tukey
correction (**l–o**) or multiple unpaired two-sided t-test with Holm–Sidak
correction (**j**). $*P < 0.05$, $***P < 0.001$, $****P < 0.0001$.
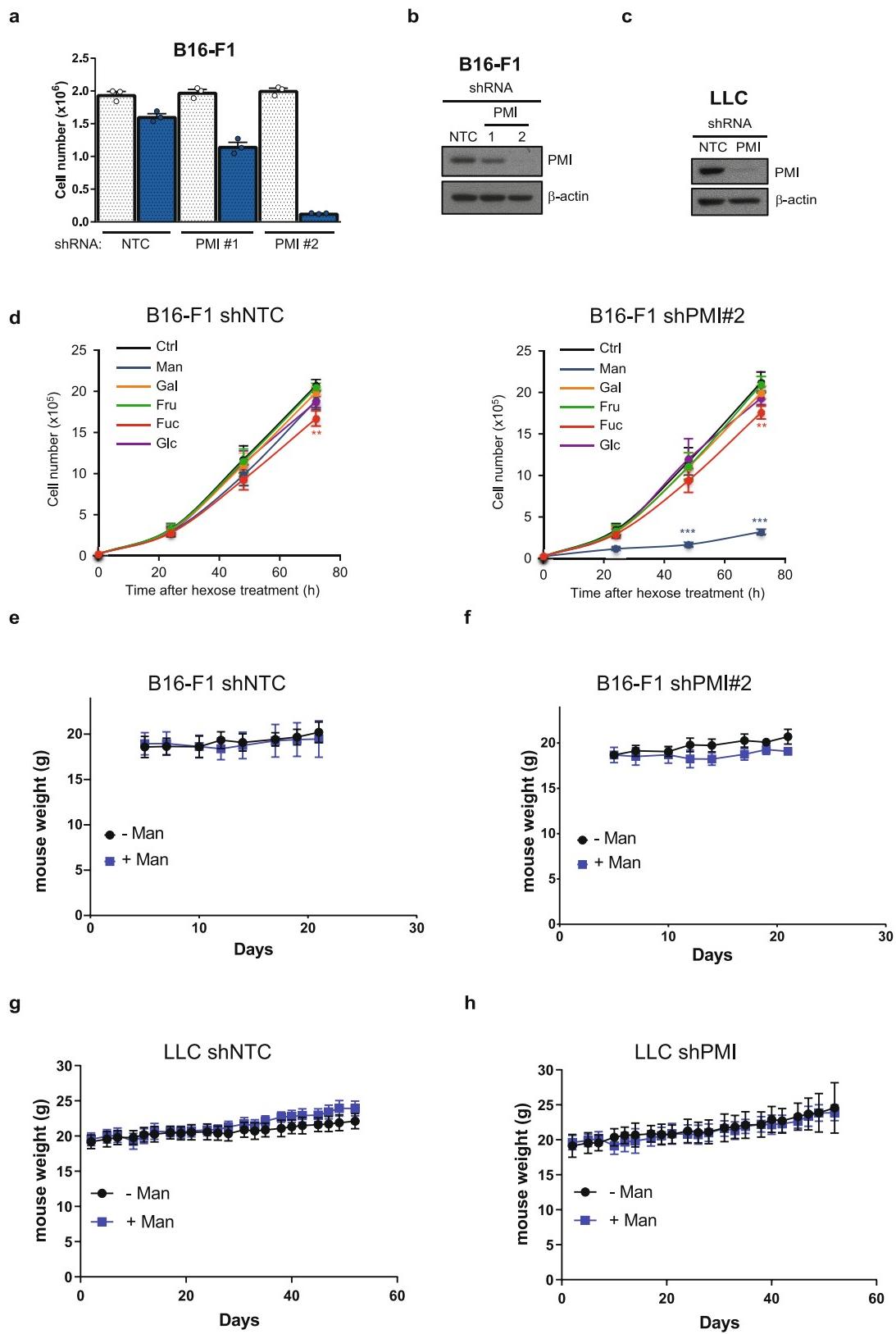
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | PMI levels dictate mannose sensitivity, and the weight of animals containing syngeneic allografts is not affected by treatment with mannose. a, b, *Mpi* knockdown causes growth retardation upon mannose treatment in a dose-dependent manner. **a,** B16-F1 cells infected with NTC or shRNA against *Mpi* were treated with (blue columns) or without (white columns) 25 mM mannose for 72 h, after which the number of cells was determined. Data are mean ± s.e.m. *n* = 3 independent experiments. **b,** The reduction in PMI was confirmed by western blotting. **c,** Western blot showing a reduction in PMI after *Mpi* knockdown in LLC cells. The data are relative to the experiments shown in Fig. 4g, j, k. **d,** B16-F1 cells infected with NTC or shRNA against *Mpi*

were treated with 25 mM mannose, 25 mM glucose, 25 mM galactose, 25 mM fructose or 25 mM fucose. At 24, 48 and 72 h after hexose treatment, the number of cells was determined. Data are mean ± s.e.m. **e–h,** Weights of mice injected with syngeneic cell lines: B16 shNTC (**e**), B16 shPMI (**f**), LLC shNTC (**g**) and LLC shPMI (**h**); the mice were given normal drinking water or drinking water supplemented with 20% mannose (*n* = 10 mice per group). All data are mean ± s.d. unless otherwise stated. In **d**, *n* = 4 independent experiments were analysed by two-way ANOVA followed by Tukey's multiple comparisons. The data in **b** and **c** were performed only once. ***$P < 0.001$.

a

PMI expression

| Negative | Low | High |

Ovarian

Breast

Prostate

Colorectal

Renal

b

Low Expression
High Expression
HR 1.13 95% CI 0.87-1.48, p=0.364

% Survival — Cancer Specific Survival (years)

c

Low Expression
High Expression
HR 0.86 95% CI 0.54-1.36, p=0.507

% Survival — Cancer Specific Survival (years)

d
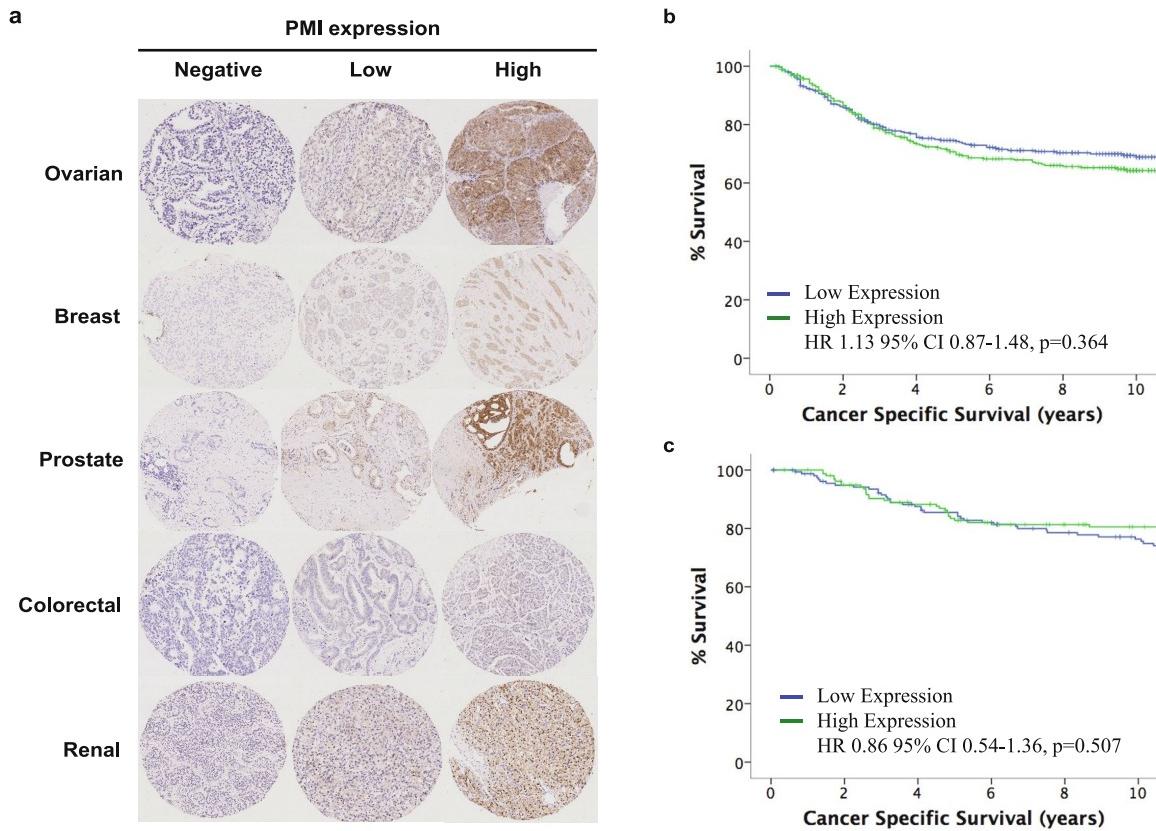
	Colorectal Cancer (n=698)			Breast Cancer (n=316)		
	N (%)	10yr CSS (SE)	*P*	*N (%)*	10yr CSS (SE)	*P*
Cytoplasmic PMI (n=698)			0.364			0.507
Low expression	347 (50)	69 (3)		157 (50)	76 (4)	
High expression	351 (50)	64 (3)		159 (50)	81 (3)	

e

AOM/DSS

mouse weight (g) — Days

f

CRC GEMM

- Man
+ Man

mouse weight (g) — Days
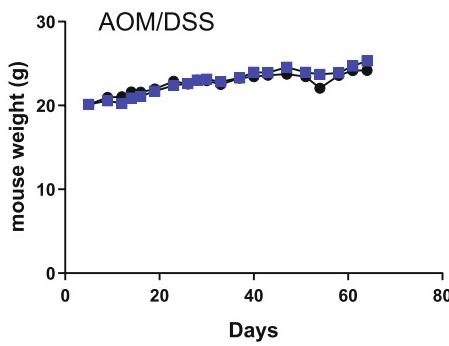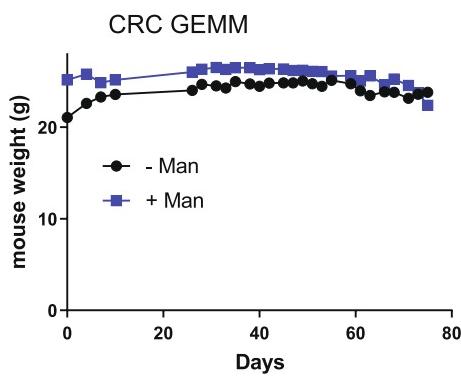
Extended Data Fig. 9 | PMI levels vary in human tumours, but do not predict overall survival and the weight of mice is not affected by mannose treatment in models of colorectal cancer. a, Images of PMI expression for each TMA, representing one negative sample (left), one low expression (middle) and one high expression (right). One sample came from each of the ovarian, breast, prostate, colorectal or renal TMAs as indicated. **b**, **c**, Kaplan–Meier curves showing cancer-specific survival based on PMI levels in $n = 698$ patients with stage I–IV colorectal cancer (**b**) or $n = 316$ patients with primary operable breast cancer (**c**). **d**, Table showing the association of PMI and cancer-specific survival. Histoscores were split into high and low expression using the ROC curve analysis for each tumour type. log-rank analysis (two-sided) was used to compare PMI and cancer-specific survival using SPSS (version 22). **e**, Mean weight of 28 mice during treatment with AOM and DSS, with or without additional mannose treatment. **f**, Weight of aged *Villin*^{creER}*Apc*$^{fl/+}$*Kras*$^{G12D/+}$ mice during treatment with or without mannose. Data are mean of each group ($n = 7$ mice, −mannose; $n = 8$ mice, +mannose).
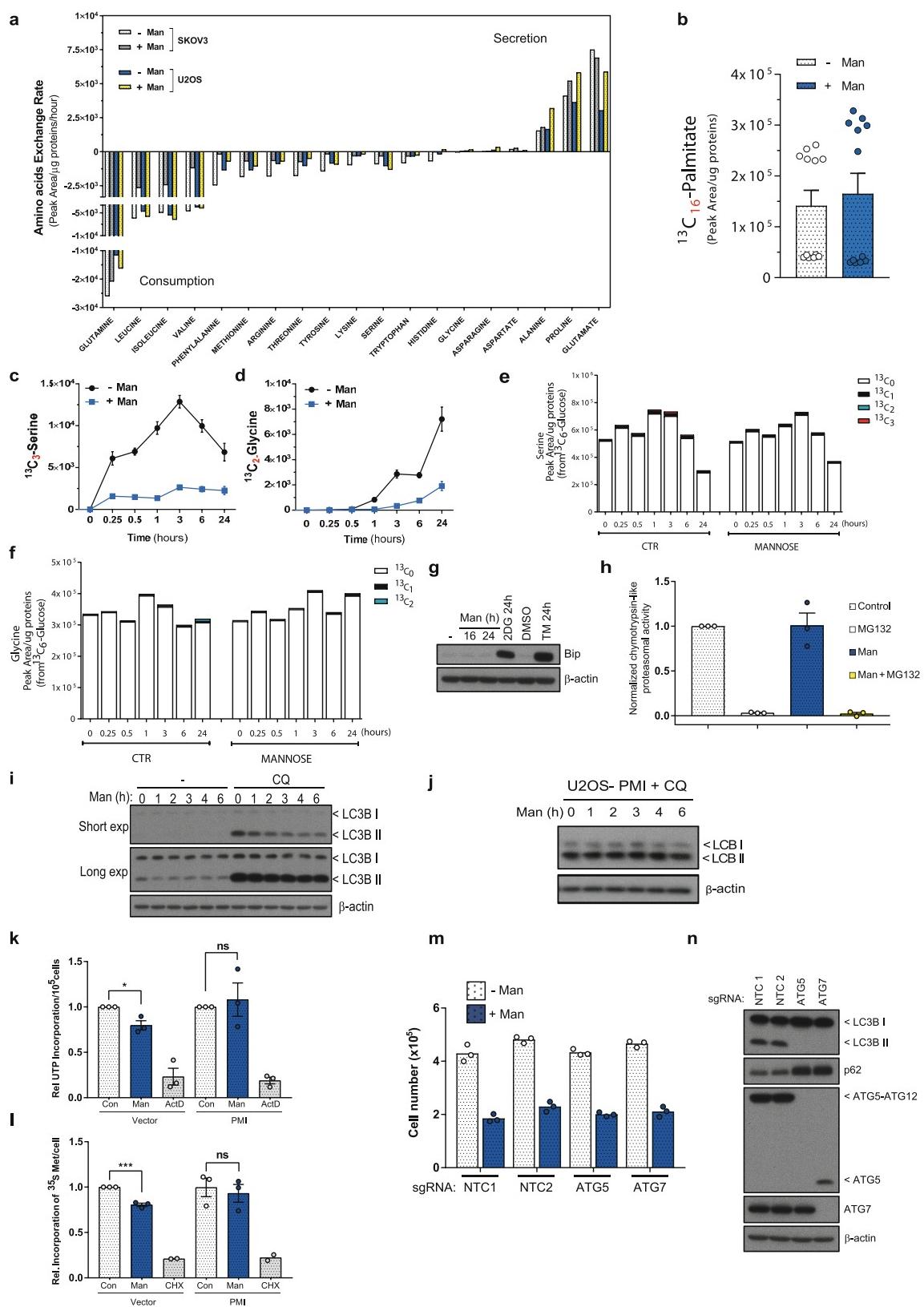
Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Mannose does not affect amino acid and fatty acid uptake, nor does it significantly affect serine and glycine levels, endoplasmic reticulum stress or proteasome activity, but it does affect autophagy, transcription and translation in a PMI-dependent manner. a, Exchange rates of amino acids between the indicated cells and their media measured after 48 h treatment with or without 25 mM mannose. Results are presented as peak area per microgram of protein per hour and are representative of one experiment. Shown is the mean of six technical replicates. **b**, Levels of $^{13}C_{16}$-palmitate (expressed as peak area per microgram of protein) in U2OS-E1a cells incubated with 50 μM $^{13}C_{16}$-palmitate, conjugated with 10% fatty-acid-free BSA, for 24 h in the presence or absence of 25 mM mannose in the culture medium (DMEM). $n = 2$ independent experiments (each involving six technical replicates). Data are mean ± s.e.m. **c**, **d**, Levels of $^{13}C_3$-serine (**c**) and $^{13}C_2$-glycine (**d**) (expressed as peak area per micrograms of protein) in U2OS-E1a cells incubated with 25 mM $^{13}C_6$-glucose for the indicated time in the presence or absence of 25 mM mannose in the culture medium (DMEM). $n = 3$ independent experiments. **e**, **f**, Distribution of isotopologues of intracellular serine (**e**) and glycine (**f**) in U2OS-E1a cells incubated with 25 mM $^{13}C_6$-glucose for 24 h in the presence (mannose) or absence (ctr) of 25 mM mannose in the culture medium (DMEM). $n = 3$ independent experiments. **g**, Mannose has no effect on the unfolded protein response. U2OS-E1a cells were treated with 25 mM mannose for 16 and 24 h. 2-Deoxy-ᴅ-glucose (2DG) and tunicamycin (TM) serve as positive controls. Induction of Bip (also known as GRP78) is a read-out of endoplasmic reticulum stress. The data are representative of three independent experiments. **h**, Proteasome activity is not affected by mannose. Mannose-sensitive U2OS-E1a cells were incubated in either DMEM or DMEM containing 25 mM mannose before measurement of proteasome activities. Cells were also treated with 10 μM MG132 as a control for proteasome activity. $n = 3$ independent experiments. **i**, **j**, U2OS-E1a cells (**i**) or U2OS-E1a (U2OS) cells overexpressing PMI (**j**) were treated with 25 mM mannose for the indicated times in the presence or absence of 20 μM chloroquine (CQ) (4 h). Western blotting was undertaken to detect LC3-B and actin. The data are representative of three independent experiments. **k**, **l**, Relative incorporation of ^{32}P UTP (**k**) or ^{35}S methionine (**l**) into U2OS-E1a control cells (vector) or U2OS-E1a cells overexpressing PMI (PMI) in the presence or absence of 25 mM mannose. Where indicated, 5 μM actinomycin D was used to inhibit transcription or 100 μg ml^{-1} cycloheximide was used to inhibit translation. In **k**, $n = 3$ independent experiments. In **l**, $n = 3$ independent experiments (control and mannose) and $n = 2$ independent experiments (CHX). **m**, **n**, U2OS-E1a cells infected with lentiCRISPR-NTC 1, NTC 2, ATG5 or ATG7 were treated with 25 mM ᴅ-(+)-mannose for 72 h. **m**, The number of cells counted after the 72-h treatment; the results represent the mean of one independent experiment performed in triplicate. **n**, Western blots show loss of LC3 lipidation and p62 accumulation in *ATG5* and *ATG7* knockout cells. Data were analysed by two-tailed unpaired *t*-tests. *$P < 0.05$, ***$P < 0.001$. Unless otherwise stated, data are mean ± s.e.m.

CAREERS

KATHRYN SAEB-PARSY

Kourosh Saeb-Parsy with his children, Nadia and Kiana, at a Mini Mudder obstacle course.

WORK–LIFE BALANCE

Fathers in science

Five scientist dads describe how they and their partners juggle their family and careers.

GERAINT REES
Be pragmatic and flexible

Dean of the faculty of life sciences at University College London. His daughter Milly is 15; his son Alex is 13.

Milly was due at the end of March 2003 but arrived in December 2002 with a 50% chance of survival, and a 50% chance of having a serious disability.

I was about to set up my own research group and achieve full independence. I'd been awarded a Wellcome Senior Clinical Fellowship and we were soon to move house.

We spent 99 days in the neonatal intensive care unit, a distressing environment.

I returned to work after three months and the focus that first year was hiring postdocs and getting back into the swing of things. Then things got more normal. We stopped worrying about Milly being rushed back into hospital. We started thinking about child care and had our first family holiday.

My wife, Rebecca, is a medical oncologist. We try to divide our child-care responsibilities 50–50. The live-out nannies we had before the children started school (a fantastic privilege to be able to afford, but also a significant cost) usually arrived around 8 a.m. and left around 6–7 p.m.. Our working pattern then had clear starts and stops. We often had to leave meetings to get home. Even now, making meetings outside normal working hours is tough. The children go to school at 8 a.m. so I can just about make 8.45 a.m.. You have to be pragmatic and flexible about what can be cancelled or rescheduled and what can't. It's important to keep talking to your partner so you don't get out of kilter with sharing child care. Otherwise you lose that feeling of shared responsibility.

Milly plays the flute and several years ago her primary school gave us six days' notice of a lunchtime concert, her first solo performance. It was also Brain Awareness Week and I was giving a lunchtime public lecture. Rebecca was in clinic.

I wanted to attend the concert, but also to deliver the lecture. I made the right decision. My colleague, Sarah-Jayne Blakemore, stepped in and her wonderful lecture was ▶

▶ much better than the one I would have given. And I got to see Milly perform.

People have occasionally said they chose my research group because of the respect I try to foster for work–life balance. I have a reputation for my out-of-office messages, which I use to communicate clearly (and humorously) that life outside work can sometimes take priority, and that family holidays are something I enjoy. I started getting congratulatory e-mails about them. Now I face pressure to come up with a funnier one whenever we go away. My last one read: "A long hot summer, epic collapse of global political discourse, impending Brexit chaos, new NSS scores, creeping REF preparation and a crescendo of short-notice UKRI deadlines. This can mean only one thing — it's time to go on vacation! I've gone with my wife and children to play, talk, swim, read, eat, drink, think, snooze and relax. And you should too — it's August, after all!"

A core message that I repeat is how much I value my family and that sometimes they will come first. But I'm also sensitive to the fact that not everybody wants to talk about their family or personal life at work.

No single solution works for everybody. I often get asked what the best time is to have children. There's never a single right time. There are trade-offs at all stages. Also, some people don't want, or can't, have children and you need to have sensitivity and respect for everyone.

At her inaugural professorial lecture, my colleague Mairéad MacSweeney talked about solidarity across genders and that men need to speak up more for flexible working. Until both men and women articulate that parenting is a normal part of working life, it won't be accepted.

Balance is a verb, not a noun. And work–life balance is not something you fix. It's something you're constantly practising and rehearsing and reflecting on.
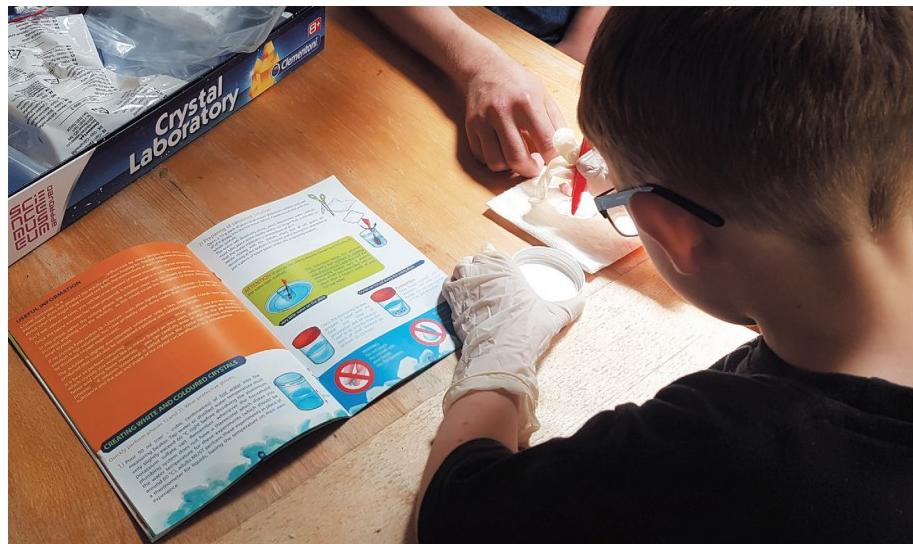
KOUROSH SAEB-PARSY
Align priorities to avoid conflict

Transplant surgeon at Cambridge University Hospitals, UK. His daughter Nadia is 8 and his daughter Kiana is 7.

My wife, Kathryn, is a primary-school teacher and has frequent evening or weekend social or work commitments. We work as a team and divide and conquer what needs to be done.

So I might pick the girls up from school or drop them off for guitar lessons because Kathryn is at work. She picks them up probably 75% of the time — but we don't keep tabs or 'scores.' If I can't pick them up from school, it is because I am doing something that Kathryn believes is good for our team.

I don't buy the idea that you have to sacrifice family life for your profession. Historically, that might have been the case, but times are changing, and I tell my students that you have to aim for both.

When there is competition between family and work, it's because everyone's priorities are not aligned. If you do start sacrificing time with your children, you're on a slippery slope that's not sustainable in the long term. I told Nadia and Kiana what I do when they were each about 3. I described what a transplant is. We've also talked about organ donation, where kidneys come from and about difficult situations in which patients have died.

I did this so that they understand why I'm not always around, usually because I'm on call. If I can't be there for lunch on Christmas Day, for example, I tell them why — saying it's because a patient has been given an amazing present, a new kidney.

They also know about my research. They love coming to the lab and meeting my group. We play (safely) with pipettes and solutions. I get them to label things. I don't want my science to be a black box to them.

I mentor people at Cambridge. We often talk about careers and family. One asked how he could finish his PhD with a second child. We discussed the importance of perspective and how, in the grand scheme of family life, work events should not always be prioritized over 'life'.

This is symptomatic of the wider issue — seeing parenting as a problem, rather than as one of the most wonderful events in your life.

We've been impressing on our daughters that they can be anything they want, and are keen to introduce them to strong female role models.

DAVID SMITH
Cement your relationships

Chemist at the University of York, UK. His son is 5.

My husband, Sam, has cystic fibrosis, and after his lung transplant in 2011, we were able to adopt. When our son arrived in 2015, I was in my early 40s and a full professor.

We wanted to split the adoption leave into six months each, but the social worker at the adoption agency preferred one primary carer for the first 12 months to help form attachments. Sam earned less than me, so he took the full year. I was disappointed and a bit jealous, but wanted to maximize my involvement and took two months of full leave and then worked at 80% of a full-time post for a year.

Many scientist mums have a hard job juggling work and family. I hope that more fathers will step up to caring responsibilities, and that this situation will change.

Our son knows that his daddy does science and wears a lab coat. He's really into volcanoes, both fascinated by and a bit scared of them.

We went to Iceland recently. When we got home, we made a plaster-cast volcano, then added bicarbonate of soda, washing up liquid and red food colouring into it. We tasted stuff for its acidity, to see what would cause an eruption. He loved that.

The chemistry department has an Athena SWAN Gold Award (a UK gender-equality charter) and meetings take place in core hours, so there shouldn't ever be one that starts before 9.45 a.m., and they should finish by 4 p.m.. We do have 9 a.m. lectures, and I have to drop my son off at school at 8.45 a.m., so I can't give those. But the department makes it easy for that to be arranged.

I used to spend around 50 nights a year away at conferences; now, it's more like 10, and never more than one night at a time. I recently did a PhD viva (oral exam) in Bristol. I left at 8 a.m. and got back at midnight, spending nine hours

David Smith's son learns to grow crystals.

DAVID K. SMITH

on a train. Previously I would have stayed over, but you have to trim those bits of academic life to enable yourself to cement relationships to support family life.

I've just turned down conference invitations to India, Canada and the United States. Employers consider business travel as part of your working hours. But it's still time away from your family.

I think the conversations that men often have at conferences can be corrosive. They talk about how big their research groups are and how many papers they have published, and not about whether they took some time off when their kids came along. That needs to change.

In academia, many people work way beyond any idea of notional contracted hours. In my experience, when you became a parent, you realize how many 'soft hours' you were putting in — those hours when you don't think you're at work. You're sitting with your partner, but you're completely absent and catching up on e-mails on your phone.

Now I try to compartmentalize work and home life, but this can lead to massive e-mail crises, especially when I am trying to do the role working 35–40 hours a week. If I spent 5 minutes on every e-mail I receive, that would be my whole working week gone.

BRIAN CAHILL
Organize your day carefully

Research programme manager at the University of Edinburgh, UK. His sons are 5 and 3.

My wife, Lini, is from Indonesia and a television producer. We were in Germany for ten years, until I moved to the University of Edinburgh in October.

We had been in Germany for five years when our first son was born. Already knowing the country and some German helped us to make some of the administration around parenting, such as registering births, easier. Also, if you're dealing with a midwife in the middle of the night, it's good to be able to speak the local language. A scientist who speaks only English and moves to a foreign country can still function as a scientist. But being able to speak the local language outside the lab bubble can help in getting things done.

Before we had kids, I would get up at 6 a.m., leave at 7 a.m., and be at work for 7.30 a.m.. When we had our first child, I still got up at 6 a.m. and mostly reached work by 8.30 a.m.. With two children, it was often around 9 a.m.. I dropped them off at kindergarten at 7.45 a.m.. It's important to be part of your children's kindergarten life (Lini picked the boys up in the afternoon). Otherwise you're excluded from a huge part of their lives.

In Germany, the working day starts at 7 a.m., so by 9 a.m. my colleagues were often having coffee. One thing we had there, which I haven't seen anywhere else in academic life, is flexitime, which was very useful. My workplace strongly discouraged weekend working, which was also great.

I struggle with switching off from work. There's always someone sending important e-mails at 5.30 p.m. on Fridays. It's not good to let people down in research, but I personally can't stay up every night until midnight when I get up at 6 a.m..

I've let go of reviewing, which is a basic part of being a researcher. You say you will do it, but it's often impossible to find the time. Writing papers and project proposals are significant for your career, but many aren't done during work time because of meetings, so you often do the work at home, which is harder with kids, who need our attention.

At home in Germany, we devoted a lot of time and energy to language learning, reading books and singing songs. Otherwise, our children would not pick up English or Indonesian. This did put a lot of responsibility on the kindergarten to teach them German, which they picked up alongside English and Indonesian.

The parental leave we got in Germany gave us 14 months between each parent. The maximum that one parent can take is 12 months. But you can also take it flexibly and extend it over 2.5 years by taking half days, for example.

Lini's parents are in Indonesia, and mine are in Ireland. My kids knew a lot of older people in the housing cooperative where we lived in Göttingen and had a grandparent-like relationship with them. There was a real sense of community there.

It's not completely impossible to be a good parent and a good researcher, but you have to organize your day.

PAUL MARTIN
Become more efficient

Cell biologist at the University of Bristol, UK. His daughter Matilda (Tilly) is 21, and his daughter Charlotte is 15.

When I mentioned that I was being interviewed for this article, some colleagues said that men don't need an article like this, that it's tougher for women. There is no doubt that my wife's career was slowed down more than mine was when the children were younger. If they were off sick from nursery, they tended to want their mum more than they wanted me, for example. Kate is also a cell biologist and head of the school of biochemistry at Bristol and, as the kids have got older, she has been able to move back up a gear with her career.

When Kate was having Tilly, a friend told her that you publish one less paper per child per year. There's some truth in that. In the early days, we shared child-care responsibility for Tilly in the evenings, although it worked out that I did two evenings and Kate did three, so there was an imbalance. If one of my child-care evenings clashed with a meeting of the London Fly Club (part of the *Drosophila* research community), I'd take Tilly along. She was about 3 or 4 at the time.

I recently came across papers that I had given Tilly one evening during a club meeting. They were subdivided into squares and she'd drawn pictures of the speakers' slides in them. One was of a fly. Another was of a gel. It was a lovely memory, but it made me a little sad. I thought: "What kind of awful parent was I?" In my opinion, we cannot be decent parents if we work until 9 p.m. I remember Kate sometimes calling and telling me to come home. But when you're setting up a lab, you sometimes have to work long hours. Science is highly competitive. As a parent, you somehow have to become more efficient.

Some colleagues have trimmed down important things such as tea-room discussions with potential collaborators, and miss seminars on areas that are slightly outside their remit.

Just working at a university puts pressure on your kids. But there are positives. Tilly recently graduated with a biology degree. As a student, she would check things with us, not the details of lectures but rather the process, and what you need, say, to get a 2:1 (a high honours degree classification in the United Kingdom, one down from a first).

One perk of being a scientist is that you can introduce your kids to useful and interesting people. Tilly just dropped by to meet one of my PhD students, a dentist. She's interested in that as a career.

Our daughters are impressed if something we do relates to their world. Earlier this year, for instance, a paper I co-authored in *Developmental Cell* was featured in *The New York Times* and received a lot of Twitter action.

For a time, I definitely was not as competitive as colleagues who did not have children. I feel that I have a more worldly perspective now — that I'm not engrossed in science all the time.

What forces me to step back (and also to realize my ignorance) is having a daughter who is revising for her biology GCSE (exams taken by secondary-school students aged 16 in England, Wales and Northern Ireland) and asks me about the difference between xylem and phloem. I tell her I know nothing about plants, although I do know lots about inflammation. She says: "You're supposed to be a professor, dad." I say: "Let's look it up." ∎

SO, ONE OF THOSE TINY ALIEN SPACESHIPS HAS FLOWN INTO YOUR HOUSE. NOW WHAT?

A self-help guide.

BY LAURA PEARLMAN

The best strategy is not to let them in, of course, but maybe you left your chimney flue open or didn't notice a hole in a window screen, and now there's a tiny alien spaceship flitting around your living room or kitchen or swooping under your dining-room table, and there's really no point in self-recrimination or speculation about which of your kids might have propped the front door open while they all loaded their cars up after Thanksgiving weekend, leaving you to rattle around alone in your empty nest once again.

However it happened, it's just you in the house with your two cats and something that looks like a miniature flying neon-pink Roomba. The first thing you need to do is take three deep calming breaths, unless of course your cat is making that weird trilling noise they make when they see a bird, in which case the first thing you need to do is get your cat the hell out of there before it tries to pounce and the aliens blast it with one of their energy cannons.

Once you've got your cat safely shut in another room — maybe your bedroom, where your other cat is already cowering under the bed — and cleaned up any scratches you might have acquired in the process, you can go back to step one. Take three deep breaths and remind yourself that the vast majority of people who have these encounters survive, usually with only minimal injuries.

Try to stay out of the aliens' way. Let them take whatever they want. No one knows for sure what motivates them, but the prevailing theory is that they're on some kind of recreational scavenger hunt, collecting a seemingly random assortment of items such as rubber bands, peanuts, crayons, lipstick, toenail clippings, blood samples and tears.

Don't fall into paranoid speculation about what the aliens are going to do with these items. They're aliens, not witches.

If you hear yowling or snarling coming from your bedroom, and you begin to regret confining your agitated cat and your terrified cat in the same room, undoing all

the progress they've made in learning to get along together over the past three years — just ignore it. You need to focus on the aliens.

Watch the ship and follow its lead. If it's repeatedly bumping against an interior door, cabinet or drawer, it wants what's inside, and your best course of action is to open the door or cabinet or drawer or whatever before the aliens get impatient and reduce your collection of stemware to a puddle of molten glass, including those nice champagne flutes you were planning to use just as soon as you had something to celebrate.

If you have a fire extinguisher, keep it to hand as you follow the aliens around the house.

If the ship is hovering near your eyes, remove your contact lenses immediately and place them on the nearest table or countertop. Seriously. You don't want the aliens to extract

them directly. If the ship is hovering near any other part of your body, remain calm and let them harvest whatever biosample they need. Chances are, it'll be something small.

Apply direct pressure to the incision site(s). If the bleeding continues for more than five minutes, call 911, and an ambulance will be dispatched to your location. Be sure to mention that aliens are present and stay on the line until they leave. The aliens will incinerate anyone attempting to enter or exit, so it's important that the paramedics wait outside until the aliens are gone.

On a related note, keep all your exterior doors and windows closed and locked.

If the ship wants to enter the room in which you've confined your cats, you really have no choice but to let it in. Your cats should be terrified enough by now that they'll stay out of the aliens' way, and as long as they do, they'll be fine. The aliens seem to be completely uninterested in cats, which is one way we know they are aliens.

Eventually, the ship will hover near an exterior door or window. This means the aliens have collected everything they set out to collect, and all you need to do at this point is let them out.

Congratulations! You've survived your first alien encounter. Extinguish any smouldering fires. If you've called 911, let them know it's safe for the paramedics to come in. Give your cats some treats. Blot any bloodstained areas of carpet with a clean, dry cloth, then apply some club soda and blot some more. Text those friends you've been meaning to get together with, and pop over to the store to pick up a bottle of champagne.

Just remember to close the door behind you on your way out. ∎

Laura Pearlman's work has appeared in Shimmer, Flash Fiction Online, Intergalactic Medicine Show *and a handful of other places. You can find her on Twitter at @laurasbadideas.*

ILLUSTRATION BY JACEY